Dcik

e 406

Vol. 45 JANUARY, 1954 No. 1

The Journal of Educational Psychology

UNIVERSITY OF MICHIGAN
MAR 3 1954
PERIODICAL READING ROOM

CONTENTS

Some Increasing Parent-Child Similarities during the Growth of Children 1
NANCY BAYLEY

Two Methods of Teaching English Composition to Retarded College Freshmen 22
RAY CHARLES MAIZE

Differential Responses among College Students on the Vocabulary Subtest of the Wechsler Intelligence Scale 29
CHALMERS L. STACEY AND S. WILLIAM SPANIER

Item Validity and Response Change under Two Different Testing Conditions 36
DOUGLAS G. SCHULTZ

Influence of Social Class upon Performance on the Draw-a-Man-Test 44
JOSEPH H. BRITTON

Classroom Environment and Pupil Welfare 52
NEWELL C. KEPHART AND WILLIAM FLOYD

Book Reviews 60

Published monthly except June to September
$7.00 a year in the United States; Canada, $7.20; other countries, $7.40.
Single issues $1.20

WARWICK & YORK, INC.
BALTIMORE 2, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879.

THE JOURNAL OF Educational Psychology

Established 1910

BOARD OF EDITORS*

STEPHEN M. COREY, *Chairman*
Teachers College
Columbia University

J. B. STROUD
State University of Iowa

WM. CLARK TROW
University of Michigan

*Appointed by the American Psychological Association.

EDITORIAL CONSULTANTS

JACK W. DUNLAP
KARL J. HOLZINER
HAROLD E. JONES
H. H. REMMERS
PERCIVAL M. SYMONDS
MILES A. TINKER
ALEXANDER G. WESMAN
PAUL A. WITTY

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY is devoted primarily to the scientific study of problems of learning, teaching, and measurement of the psychological development of the individual. The JOURNAL will contain articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, and personality, particularly as related to school adjustment; emotion, motivation, and character; mental development and methods. This last will include tests, statistical techniques, and research techniques in cross-sectional and developmental studies.

INFORMATION FOR CONTRIBUTORS

Manuscripts and communications regarding editorial matters should be addressed to the Chairman of the Board of Editors, Stephen M. Corey, Teachers College, Columbia University, New York 27, N. Y.

THE JOURNAL has set regulations regarding content and style of material published, and these should be observed in the preparation of manuscripts to be submitted.

Tables and graphs.—In keeping with the practice of most professional periodicals of like character, THE JOURNAL now requires authors to bear part of the added cost resulting from the use of tables, formulas, and graphs. Prospective contributors may obtain a detailed statement as to

(Continued on Inside Back Cover)

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 45 January, 1954 Number 1

SOME INCREASING PARENT-CHILD SIMILARITIES DURING THE GROWTH OF CHILDREN

NANCY BAYLEY

Institute of Child Welfare, University of California

Age-changes in the correlation between children's mental test scores and some measure of parental ability have been shown for several different populations. Such reports have often been incidental to other aspects of research on intelligence. It has seemed worth while, therefore, to relate some of these findings to relevant material, in both mental and physical relationships, in the Berkeley Growth Study.

PARENT-CHILD CORRELATIONS IN THE LITERATURE

It is only rarely that mental test scores are available for the parents of tested children. But there are a few instances in which intelligence test scores of parents and their children have been compared directly. One of the most extensive and important of these was made by Jones and Conrad (*8*, *11*, *12*). They correlated Army Alpha test scores of parents with Alpha and Stanford-Binet scores of their children. In one hundred and forty families, in which both parents and children were given the Army Alpha test, four hundred and forty-one pairings gave a parent-child *r* of .49. The same *r* of .49 was obtained for five hundred and one pairings in one hundred eighteen families in which the children had the Stanford-Binet test. The *r* is also .49 for separate mother-child and father-child comparisons.[1]

Among the most recently published parent-child test comparisons are the correlations presented by Skodak and Skeels between

[1] See (*13*) for a report on age differences in *r*'s for this sample. In earlier studies the authors did not break down the data by comparing children of different ages with their parents.

IQ's of true mothers (tested near the time the children were born) and their children who have grown up in foster homes (*15*). In this study the *r*'s for sixty-eight cases gradually rose from zero when the children were two years of age to .38 and .44 on the latest tests of the children, given when they averaged thirteen and one-half years.

The *r*'s for these two studies are, for several reasons, not directly comparable. However, the earlier study gives the general relationships to be expected when scores are available on both parents, and for a large population of children (age three years and older) growing up with their own parents. In the later study the comparisons are made repeatedly for the same population, the variable factor being the ages of the children at the time of testing. Although in this study the *r*'s are lower, they give evidence of increasing mother-child similarity in intellectual capacities as the children grow older. This is impressive in view of the fact that the children were separated from their true mothers in early infancy. Furthermore, correlations of the children's IQ's with educational status of foster parents remained close to zero (ranging at the thirteen and one-half year tests from .00 to .06).

In the absence of direct measures of intelligence, some other estimate of parental ability has often been used. We will concern ourselves here only with some of these studies which report correlations for children of different ages. In the Berkeley Growth Study (*2*, *6*), and in the Guidance Study which has been reported by Honzik (*10*), several socio-economic factors have been correlated with the children's mental test scores at successive ages up to eight and ten years. In both studies the parent-child *r*'s are zero or even negative at first, but increase as the children grow older. Honzik found that by the time the Guidance Study children were eight years old, their mental test scores correlated .36 with mid-parent education, .41 with a socio-economic index, and .55 with a rating of parent's intelligence. By the time the Berkeley Growth Study children (*2*) were ten years old, their intelligence scores correlated around .55 with such factors as socio-economic ratings and education of parents.

PARENT-CHILD CORRELATION IN THE BERKELEY GROWTH STUDY

The Sample.—In the present report some parent-child comparisons are made on both mental and physical characteristics of the

TABLE I.—MEANS AND STANDARD DEVIATIONS OF COMPARABLE ADULT EDUCATIONAL AND PHYSICAL DATA FOR CHILDREN AND THEIR PARENTS

Heights (Inches)	N	Mean	SD
Mature* boys	22	71.63	2.38
Boys at 18 years	21	71.36	2.51
Parents of boys			
Fathers	22	70.57	2.46
Mothers	22	64.97	2.50
Mid-Parent†	22	70.37	2.11
All fathers	42	69.69	2.73
Girls at 18 or mature	21	65.29	2.57
Parents of girls			
Fathers	20	68.72	2.67
Mothers	20	63.92	2.33
Mid-Parent†	20	63.67	1.90
All mothers	42	64.47	2.48
Weights (Pounds)			
Mature boys	22	157.49	13.09
Boys at 18 Years	21	154.40	12.64
All fathers	41	161.71	19.30
Mature girls	24	129.87	19.45
Girls at 18 years	20	126.90	16.15
All mothers	42	132.76	18.74
Weight/Height^2 Indices‡			
Boys at 21 years	17	218.8	17.5
Boys at 18 years	21	213.2	12.2
All fathers	41	233.0	25.1
Girls at 21 years	20	210.1	28.6
Girls at 18 years	20	212.6	26.6
All mothers	42	224.2	27.1
Education (Years of Schooling Completed)			
Boys	23	13.87	3.38
Girls	23	14.67	2.05
Fathers	43	13.66	3.82
Mothers	43	12.79	3.12
Mid-Parent	43	13.22	2.89

* Heights after closure of all epiphyses in hand and knee, often after eighteen years.

† Boys' mothers' heights multiplied by 1.08; girls' fathers' heights multiplied by 0.923.

‡ Indices computed from metric units.

children in the Berkeley Growth Study whose records are now complete through twenty-one years. The number of children tested at any given age varies from forty to sixty-one. Data are available for eighteen years or longer for forty-seven children from forty-three families. Table I gives the means and SD's for both parents and children in the characteristics to be considered here.

In heights and weights this group of parents is probably typical of California adults of similar age and racial antecedents. They are, for the most part, better educated, and have a generally higher socio-economic status, than a cross section of the Berkeley population with which they were compared at the outset of the Growth Study (*6*).

In comparison with their parents the children seem to be favored, both physically and mentally. They are, on the average, about an inch taller than their parents, but five pounds lighter. Consequently, they are slenderer in build, but this difference may be due to the fact that they are still younger than most of the parents were at the time the parental measurements were obtained.

Even though in a number of instances the children's educations are not yet completed, we find that they have had more education than their parents. This is true of the girls, in particular, but at least four of the boys are now in medical school or other graduate study. The mean IQ of these children at eighteen years on the Wechsler-Bellevue Scale was 122, SD 16 (*3*).

Mental.—As stated above, we do not have intelligence-test scores for the parents. Among the socio-economic variables studied, the parents' education was found to correlate most highly with the children's mental-test scores.

In Table II and Figure 1 these previously reported relations for the Berkeley Growth Study are reviewed and brought up to date. For graphical representation the product-moment r's have been transformed to z's.[2] Iu this figure, the z transformations are based on r's between the midparent education (average of the years of schooling completed by mother and father) and the children's mental test sigma scores at every test age from birth through eighteen years. That is, we have compared a single evaluation of parental ability with a succession of different scores of their children's intelligence. It is evident that the parent-child correlation

[2] From the table on page 348 in McNemar (*14*).

TABLE II.—CORRELATIONS (AND Z SCORES) BETWEEN MID-PARENT EDUCATION AND CHILD'S MENTAL TEST SCORES AT EACH AGE TESTED

Age		Mid-Parent Education by	*N*	*r*	*z*
Mo.	1	California First-Year Scale	52	−.01	−.01
	2	California First-Year Scale	58	−.02	−.02
	3	California First-Year Scale	61	−.27*	−.28
	4	California First-Year Scale	58	−.22	−.22
	5	California First-Year Scale	58	−.29*	−.30
	6	California First-Year Scale	57	−.23	−.23
	7	California First-Year Scale	52	−.19	−.19
	8	California First-Year Scale	53	−.05	−.05
	9	California First-Year Scale	56	+.06	+.06
	10	California First-Year Scale	56	+.08	+.08
	11	California First-Year Scale	52	+.00	+.00
	12	California First-Year Scale	53	+.03	+.03
	13	California First-Year Scale	53	−.05	−.05
	14	California First-Year Scale	46	−.02	−.02
	15	California First-Year Scale	52	−.00	−.00
	18	California Preschool Scale	51	+.16	+.16
	21	California Preschool Scale	53	.29†	.30
	24	California Preschool Scale	48	.50†	.55
	27	California Preschool Scale	51	.41	.44
	30	California Preschool Scale	47	.44	.47
	36	California Preschool Scale	49	.47	.51
	42	California Preschool Scale	43	.35	.37
	48	California Preschool Scale	45	.50	.55
	54	California Preschool Scale	44	.59	.68
	60	California Preschool Scale	47	.58	.66
Yr.	6	1916 Stanford-Binet	48	.57	.65
	7	1916 Stanford-Binet	46	.50	.55
	8	1937 Stanford-Binet	47	.59	.68
	9	1937 Stanford-Binet	45	.57	.65
	10	1937 Stanford-Binet	47	.54	.60
	11	1937 Stanford-Binet	48	.61	.71
	12	1937 Stanford-Binet	46	.59	.68
	13	Terman-McNemar	41	.62	.72
	14	1937 Stanford-Binet	40	.60	.69
	15	Terman-McNemar	42	.56	.63

TABLE II.—*Continued*

Age	Mid-Parent Education by	N	*r*	*z*
16	Wechsler-Bellevue	44	.64	.76
17	1937 Stanford-Binet	43	.65	.78
18	Wechsler-Bellevue	40	.60	.69

* The negative *r*'s for months three and five are significant at the .05 level of confidence or better.

† The positive *r* at month twenty-one is at the .04 level of confidence. After this the *r*'s are all highly significant, usually above .001.

is at first zero and even becomes negative during the second quarter year (at month five the *r* of −.29 is significant at the .03 level of confidence). The *r*'s hover around zero from seven months through fifteen after which they rise rapidly to .50 at two years.

The positive *r*'s become statistically significant after eighteen months, the level of confidence being .01 or better from two years on.

When the sexes are treated separately, there appears to be little sex difference in the parent-child relationships in mental ability, in this group, although the girls' scores are oftener than the boys' significantly correlated with parental status. The *r*'s are given for ten ages in Table III.

Bayley and Jones, in an early report on these changing relationships (*6*), discussed the question of the probable contributions of hereditary and environmental factors. The evidence seemed to indicate that hereditary factors emerged gradually as the children matured, but the proportional contribution of environmental factors could not be determined. Conrad in 1931 (*7*) reviewed studies of kin resemblances in physique and intelligence. He concluded that when various relevant factors were controlled, the kin resemblances in physical and intellectual characteristics were about the same. He pointed out, however, that the mere similarity of correlation coefficients gives no proof of similar etiology of the resemblances.

Physical.—The question has arisen whether changes in parent-child similarities might occur during growth in other areas as well as in mental scores. Are there similar changes, for example, in physical characteristics which are relatively unaffected by environmental differences, given a sample of normally healthy children (*1*) within a normal environmental range? To test this, we have

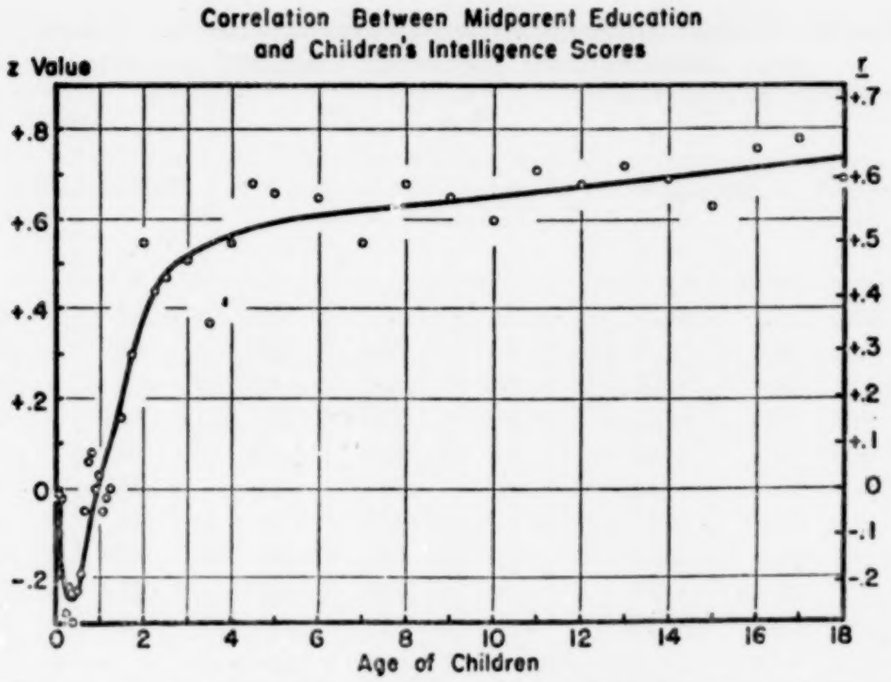


FIG. 1. Correlations between Children's Intelligence Sigma Scores and the Years of Schooling (Average of Mother and Father) of the Parents are Plotted by Age. Z-values are shown at the left, Correlation Coefficients at the right.

correlated some available parent-child physical data, i.e., heights and weights, for the Berkeley Growth Study.

The parents' heights and weights were obtained by the visiting nurse in an interview which occurred soon after the children were born, at a time when most of the parents were young adults in their twenties. The measurements were not made under laboratory conditions, but may be considered fairly good approximations. The children were measured nude, during their regular visits to the Institute, according to exact anthropometric techniques (*5*).

Correlations have been computed for three physical variables (height, weight, and an index of build, W/H^2), for the parents as adults with these same measurements of the children made at nineteen ages (and in some cases thirty ages). These *r*'s are given in Tables IV through VII. Because of physical sex differences, it is necessary to treat the sexes separately.

The parent-child correlations in height are given in Table IV. The *r*'s with sons tend to increase with age, although unevenly. The father-son *r*'s are significant after three years, but the mother-son *r*'s are not significant until thirteen years. The girls are significantly correlated with their mothers after the first few months, but not with their fathers, except possibly toward maturity.

TABLE III.—CORRELATIONS BETWEEN CHILDREN'S MENTAL TEST SCORES AND THEIR PARENTS' EDUCATION: SEXES SEPARATE

CA	N	Boys		N	Girls	
		Father	Mother		Father	Mother
		r	*r*		*r*	*r*
Mo. 6	31	−.29	.00	29	−.14	−.05
Yr. 1	28	.04	.24	28	−.08	.02
3	26	.39*	.49†	26	.26	.31
6	25	.34	.47*	26	.56†	.61†
8	24	.34	.40*	26	.64†	.59†
10	24	.44*	.38	26	.58†	.51†
14	21	.45*	.45*	21	.60†	.52*
15	21	.42	.42	21	.52*	.42
16	22	.48*	.40	22	.61†	.63†
18	21	.43*	.41	19	.60†	.63†

* Significant at the .05 level of confidence or better.
† Significant at the 0.1 level of confidence or better.

Midparent values for height have been computed by making an adjustment in the height of the parent of the sex opposite to that of the child (*9*). This is done by multiplying the heights of the boys' mothers by 1.08 and the heights of the girls' fathers by 0.923 and averaging the adjusted heights of the two parents. (In this sample the boys' mothers and fathers correlate with each other .31 in height and .04 in weight. For the girls' parents the corresponding *r*'s are: height, .10, weight .11). The correlations of the children's heights with the midparent values are shown in Table V and Figure 2. The boys show a continuously increasing correlation with their parents; the *r*'s become significant at about four years of age. The girls show an early rapid increase, becoming highly significant by five months, after which the *r*'s level off, then seem to pick up again after twelve years. However, from the data in Table IV, it would appear that these relationships with the midparent values are due primarily to similarities with the parents of the same sex.

The correlations for weight are shown in Table VI. The parent-son *r*'s are rather low, but there is a trend toward increasing relationship between the boys and their fathers, the *r*'s reaching the .05 level of confidence by fifteen years. The father-daughter *r*'s indicate some relation by five years, but they never reach statistical

TABLE IV.—CORRELATIONS (AND Z SCORES) BETWEEN HEIGHTS OF PARENTS AND THEIR CHILDREN

Age	Boys					Girls				
	N	Fathers		Mothers		N	Fathers		Mothers	
		r	z	*r*	z		*r*	z	*r*	z
Mo. 1	24	.30	.31	.05	.05	26	.36	.38	.30	.31
6	30	.25	.26	.27	.28	28	.42*	.45	.47†	.51
Yr. 1	28	.30	.31	.31	.32	28	.36	.38	.67†	.81
2	23	.28	.29	.21	.21	19	.34	.35	.56*	.63
3	25	.45*	.49	.25	.26	24	.38	.40	.67†	.81
4	22	.52*	.58	.25	.26	23	.35	.37	.66†	.79
5	23	.52*	.58	.30	.31	24	.24	.25	.70†	.87
6	24	.52†	.58	.35	.37	24	.26	.27	.75†	.97
7	24	.53†	.59	.35	.37	21	.33	.34	.76†	.996
8	23	.54†	.60	.43	.46	24	.26	.27	.74†	.95
9	23	.56†	.63	.43	.46	22	.30	.31	.71†	.89
10	23	.60†	.69	.37	.39	24	.24	.25	.72†	.91
11	24	.57†	.65	.41*	.44	21	.29	.30	.68†	.83
12	20	.53*	.59	.41	.44	22	.24	.25	.74†	.95
13	24	.45*	.49	.50*	.55	17	.56*	.63	.62†	.73
14	20	.58†	.66	.69†	.85	17	.49*	.54	.66†	.79
15	21	.42	.45	.71†	.89	20	.43	.46	.68†	.83
16	22	.47*	.51	.55†	.62	19	.42	.45	.66†	.79
17	20	.64†	.76	.60†	.69	20	.44	.47	.66†	.79
18	21	.52*	.58	.44	.47	19	.64*	.76	.52†	.58
21	17	.44	.47	.38	.40	20	.50*	.55	.67†	.81

* Significant at the .05 level of confidence or better.
† Significant at the .01 level of confidence or better.

significance. The mother-daughter correlations, however, show strongly increasing similarities in weight through sixteen years. They reach the .05 level of confidence with an *r* of .43 at two years and the .01 level by six years, when the *r* is .57, with *r*'s considerably higher at most of the later ages.

In our studies of growth we have found the W/H^2 index very useful as a measure of body build in children. It gives a measure of breadth, or volume, relative to stature. The parent-child correlations for this index are given in Table VII. The relative independence of the boys, especially in regard to weights, is reflected in

TABLE V.—CORRELATIONS BETWEEN MID-PARENT HEIGHTS AND HEIGHTS OF THEIR CHILDREN

Age	N	Boys		N	Girls	
		r	*z*		*r*	*z*
Mo. 1	24	.21	.21	25	.46*	.50
2	32	.22	.22	27	.41*	.44
3	32	.24	.24	30	.42*	.45
4	31	.27	.28	29	.58†	.66
5	31	.23	.23	27	.56†	.63
6	32	.35*	.37	29	.61†	.71
7	27	.37	.39	24	.60†	.69
8	26	.36	.38	28	.48†	.52
9	28	.36*	.38	29	.67†	.81
10	29	.32	.33	27	.63†	.74
11	28	.36	.38	23	.51†	.56
12	29	.39*	.41	26	.67†	.81
18	23	.22	.22	23	.64†	.76
Yr. 2	27	.36	.38	23	.66†	.79
3	26	.44*	.47	23	.69†	.85
4	25	.52†	.58	23	.60†	.69
5	23	.50*	.55	24	.60†	.69
6	26	.56†	.63	24	.61†	.71
7	24	.54†	.60	22	.61†	.71
8	24	.58†	.66	24	.59†	.68
9	24	.60†	.69	22	.61†	.71
10	24	.60†	.69	24	.59†	.68
11	25	.60†	.69	21	.58†	.66
12	21	.59†	.68	22	.62†	.72
13	24	.62†	.72	17	.69†	.85
14	20	.82†	1.16	17	.69†	.85
15	21	.68†	.83	20	.67†	.81
16	22	.62†	.72	19	.67†	.81
17	20	.77†	1.02	20	.67†	.81
18	21	.58†	.66	17	.76†	1.00
21	17	.51*	.56	20	.69†	.85

* Significant at the .05 level of confidence or better.

† Significant at the .01 level of confidence or better.

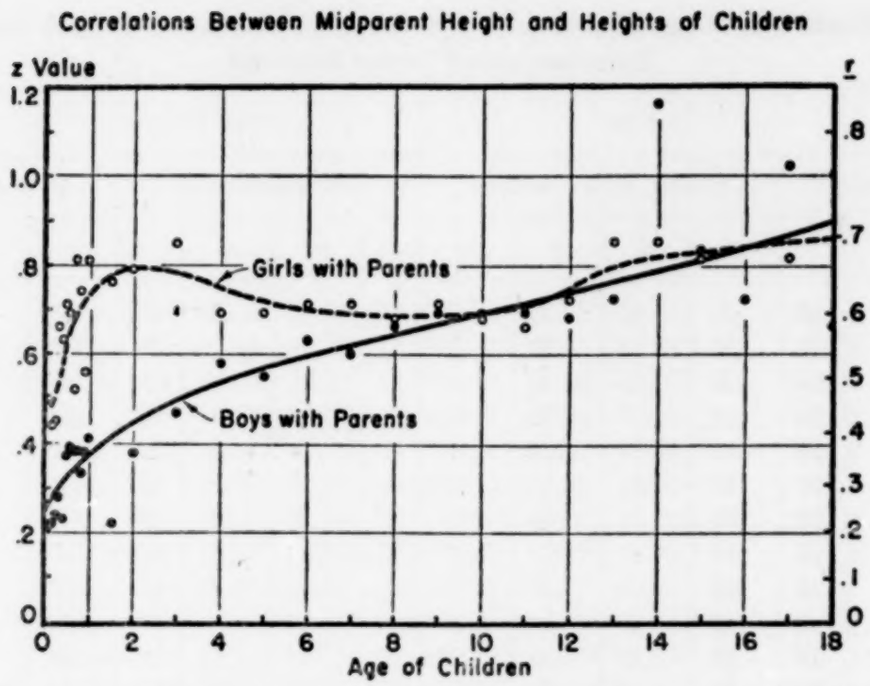


FIG. 2. The Solid Line and Dots Show the Z-values (and Correlations) between the Heights of the Boys and a Midparent Score (Boys' Mothers' Heights are Multiplied by 1.08 and Averaged with the Fathers' Heights). Broken Line and Open Dots Show these Values for the Girls and their Parents (Girls' Fathers' Heights are Multiplied by .923 before averaging with the Mothers' Heights).

these r's: in the W/H^2 index the boys are unrelated to their parents at all ages. The girls' indices however, show increasing r's with both parents. The father-daughter r's approach significance at years eight, nine and ten, and again at sixteen and seventeen years. Mother-daughter r's are usually above the .01 level of confidence at most of the later ages.

The sex differences in correlation are in some instances very striking. Even though the samples are small, statistical tests indicate that the differences may well be significant (Table VIII). In heights (Tables III and VIII) the boys are found to be more closely related to their fathers than to their mothers, at the .05 level of confidence or better, at years four, five, six, ten, and eleven; while the girls are more closely related (.05 level) to their mothers at all ages except thirteen and eighteen years, the difference being highly significant for years seven to twelve (.01 or better).

In general, also, the girls show more significant relation to their parents than do the boys. Their height correlations with midparent

TABLE VI.—CORRELATIONS (AND Z SCORES) BETWEEN WEIGHTS OF PARENTS AND THEIR CHILDREN

Age	Boys					Girls				
	N	Fathers		Mothers		N	Fathers		N	Mothers	
		r	*z*	*r*	*z*		*r*	*z*		*r*	*z*
Mo. 1	25	−.05	−.05	.35	.37	25	−.20	−.20	26	.41*	.44
6	31	.13	.13	.23	.23	27	−.16	−.16	28	.23	.23
Yr. 1	28	.19	.19	.20	.20	27	.12	.12	28	.35	.37
2	25	.26	.27	.02	.02	21	.01	.01	22	.43*	.46
3	25	.33	.34	.07	.07	23	.24	.25	24	.47*	.51
4	21	.23	.23	−.12	−.12	22	.19	.19	23	.41*	.44
5	23	.34	.35	.09	.09	23	.37	.39	24	.47*	.51
6	24	.32	.33	.23	.23	23	.30	.31	24	.57†	.65
7	24	.23	.23	.20	.20	20	.45*	.49	21	.62†	.73
8	23	.29	.30	.10	.10	22	.40	.42	23	.67†	.81
9	23	.28	.29	.06	.06	20	.40	.42	21	.68†	.83
10	23	.31	.32	−.01	−.01	23	.30	.31	24	.74†	.95
11	24	.23	.23	.07	.07	20	.32	.33	21	.71†	.89
12	20	.30	.31	.01	.01	21	.20	.20	22	.73†	.93
13	24	.30	.31	.17	.17	16	.19	.19	17	.51†	.56
14	20	.36	.38	.15	.15	16	.22	.22	17	.89†	1.42
15	21	.46*	.50	.15	.15	19	.25	.26	20	.88†	1.38
16	22	.45*	.49	−.09	−.09	18	.30	.31	19	.85†	1.26
17	20	.49*	.54	.28	.29	19	.42	.45	20	.67†	.81
18	21	.44	.47	.13	.13	18	.31	.32	19	.43	.46
21	17	−.02	−.02	.10	.10	19	.16	.16	20	.81†	1.13

* Significant at the .05 level of confidence or better.
† Significant at the .01 level of confidence or better.

height (Tables IV and VIII) are significant at the .05 level or better for all but two of the fifteen *r*'s during the first three years: after that the boys' *r*'s are equally high. In weight (Tables V and VIII) girls are more significantly correlated with their mothers than are boys with their fathers, at thirteen of twenty ages, eight of these at the .01 level of confidence or better.

CHILDREN'S SELF-CORRELATIONS

It may be relevant here to point to the low relationship between these children's measurements in infancy and their own measurements as young adults. From Table IX, we see that mental test

TABLE VII.—CORRELATIONS (AND Z SCORES) BETWEEN W/H² INDEX OF PARENTS AND THEIR CHILDREN

Age	Boys					Girls				
	N	Fathers		Mothers		N	Fathers		N	Mothers	
		r	z	r	z		r	z		r	z
Mo. 1	24	−.15	−.15	.32	.33	25	−.15	−.15	26	.13	.13
6	30	−.09	−.09	.23*	.23	27	−.31	−.32	28	.13	.13
Yr. 1	28	−.12	−.12	.30*	.31	27	−.10	−.10	28	.07	.07
2	24	.23	.23	−.11	−.11	19	−.04	−.04	19	.26	.27
3	25	−.04	−.04	.16	.16	23	−.01	−.01	24	.14	.14
4	21	−.27	−.28	.13	.13	22	.06	.06	23	.13	.13
5	23	−.07	−.07	.26	.27	23	.19	.19	24	.14	.14
6	24	−.24	−.25	.28	.29	23	.33	.34	24	.22	.22
7	24	−.14	−.14	.31	.32	20	.39	.41	21	.39	.41
8	23	−.18	−.18	.14	.14	22	.46*	.50	23	.44*	.47
9	24	−.12	−.12	.04	.04	20	.45*	.49	21	.53*	.59
10	23	−.04	−.04	−.16	−.16	23	.42*	.45	24	.64†	.76
11	24	−.03	−.03	−.03	−.03	20	.31	.32	21	.75†	.97
12	20	−.06	−.06	−.10	−.10	20	.37	.39	21	.65†	.78
13	24	.09	.09	.06	.06	16	.40	.42	17	.35	.37
14	19	−.08	−.08	.22	.22	16	.30	.31	17	.81†	1.13
15	20	.17	.17	.16	.16	19	.31	.32	20	.78†	1.05
16	22	−.09	−.09	.08	.08	18	.44	.47	19	.67†	.81
17	20	.14	.14	.14	.14	19	.51*	.56	20	.41*	.44
18	21	.14	.14	.03	.03	18	.47*	.51	19	.17	.17
21	17	−.54	−.60	.25	.26	19	.43	.46	20	.69†	.85

* Significant at the .05 level of confidence or better.

† Significant at the .01 level of confidence or better.

scores earned during the first year are unrelated to their eighteen-year test scores (*3*). As a matter of fact, the *r*'s of tests during the first nine months drop to zero or below when comparisons are made with scores at three years or older. However, the two-year and eighteen-years tests correlate .55, and starting at seven years all *r*'s with eighteen-year tests are above .80.

Growth in size is somewhat more stable, as may be seen in Table X. During the first year, the children's heights correlate in the neighborhood of .5 to .6 with their heights as adults. By the time they are three years old the heights of both sexes correlate above .83 with their own adult statures.

The self-correlations in weights, on the other hand, present a

TABLE VIII.—SIGNIFICANCE OF SEX DIFFERENCES IN PARENT-CHILD CORRELATIONS

	In Height				In Weight	
	Boys correlate higher with Fathers at this level of confidence*	Girls correlate higher with Mothers at this level of confidence*	Girls correlate higher with Midparent scores at this level*		Girls correlate higher with Mothers than Boys with Fathers at this level*	
Yr. 1	—	.01	Mo. 1	.03	Mo. 1	.01
2	—	.04	2	.04	Yr. 2	.05
3	—	.02	3	.04	3	—
4	.03	.02	4	.02	4	.05
5	.04	.005	5	.015	5	—
6	.05	.02	6	.02	6	.03
7	—	.005	7	.03	7	.01
8	—	.003	8	—	8	.01
9	—	.007	9	.01	9	.01
10	.01	.003	10	.015	10	.005
11	.05	.01	11	—	11	.004
12	—	.003	12	.02	12	.006
14	—	.05	18	.01	13	.05
15	—	.03	Yr. 2	.02	14	.001
16	—	.03	3	.02	15	.001
17	—	.04			16	.001
					17	.04

* Not given if below the .05 level of confidence.

picture more like that of the parent-child relationships for weight. Although the self-correlations are pretty consistently higher than the same-sex parent-child r's, the trends of correlation, with age, tend to be parallel.

Some correlations illustrating different degrees of consistency in build are shown for two indices in Table XI. The Stem Length/Height index is used to characterize the relative lengths of the legs and the body stem. The indices obtained on the Berkeley Growth Study cases at eighteen years are correlated with their indices at three earlier ages, three, twelve and thirty-six months. For both sexes the r's are similar. They are lowest at one year, and in general show only a moderate relation to later status. The Weight/Height^2 index is correlated for the same age-intervals. This index of build shows no consistency in the boys between eighteen

TABLE IX.—CORRELATIONS BETWEEN EARLIER MENTAL TEST SCORES AND WECHSLER-BELLEVUE IQ AT 18 YEARS

Age	Test	N	r
Mo. 1	California First-Year Scale	36	.07
6	California First-Year Scale	36	−.12
Yr. 1	California First-Year Scale	35	.25
2	California Preschool Scale	36	.55
3	California Preschool Scale	36	.49
4	California Preschool Scale	34	.66
5	California Preschool Scale	36	.76
6	1916 Stanford-Binet	36	.77
7	1916 Stanford-Binet	36	.81
8	1937 Stanford-Binet, L	36	.85
9	1937 Stanford-Binet, L	34	.87
10	1937 Stanford-Binet, M	36	.86
11	1937 Stanford-Binet, L	35	.93
12	1937 Stanford-Binet, M	34	.89
13	Terman-McNemar, C	31	.93
14	Stanford-Binet, L	33	.89
15	Terman-McNemar, D	33	.88
16	Wechsler-Bellevue	36	.94
17	Stanford-Binet, M	36	.90

years and the first three years of life. The girls, on the other hand, appear to be very stable from one year.

The greater stability of the physical measurements and especially of heights, could be due to any or all of the following: (a) greater reliability of measurement, (b) greater stability during growth of the component parts in the values compared, and (c) smaller influence of environmental forces in producing individual differences, at least for this sample. Weight is very reliably measured, but it does vary with growth changes in body proportions, and with health and nutrition.

INDIVIDUAL TRENDS IN GROWTH

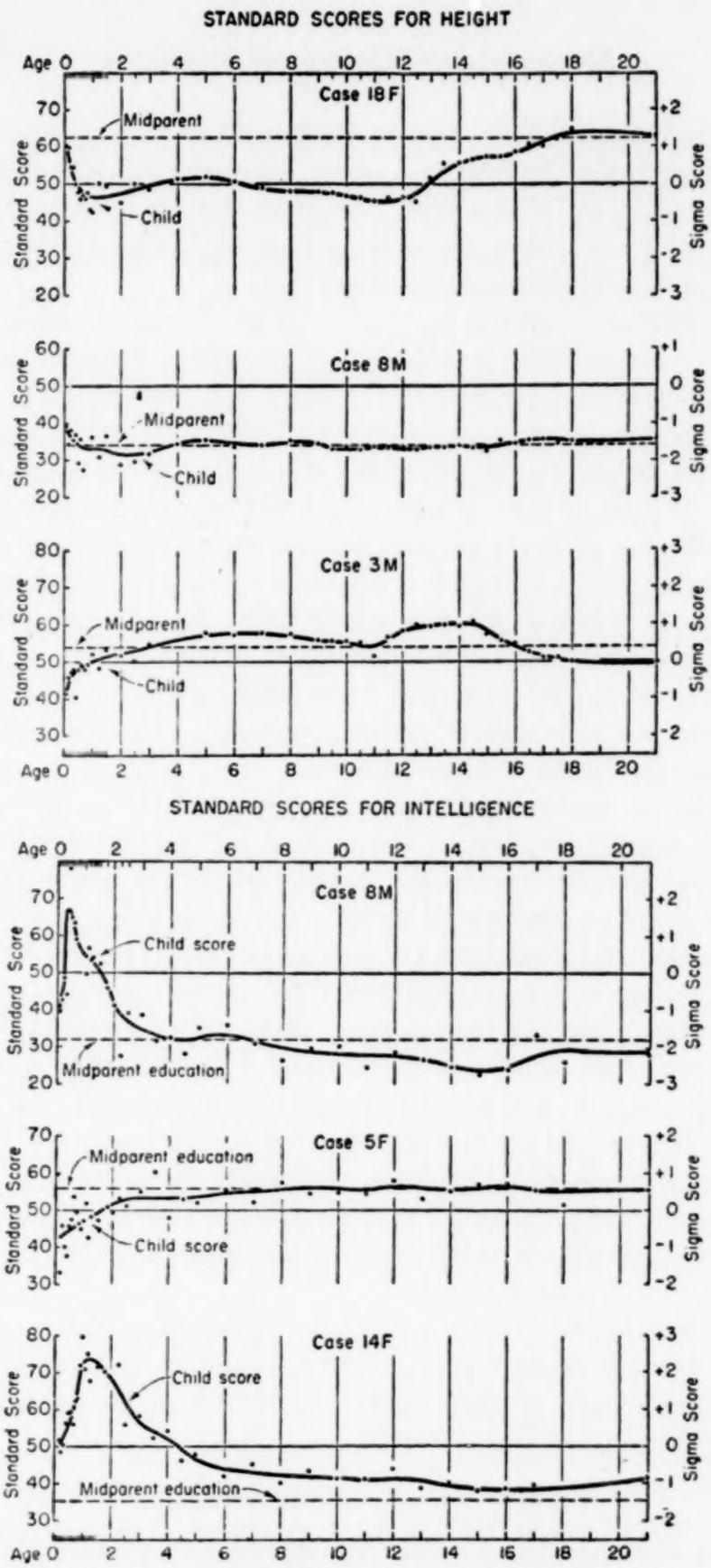
The changing status of a child, during his growth, may be illustrated by plotting a curve of his relative (or sigma) scores with age. Six such curves are shown in Figures 3 and 4. They include both mental scores and height scores, and on each child's chart the relative status of his own parents (compared with the parental

TABLE X.—CHILDREN'S EARLIER HEIGHTS AND WEIGHTS CORRELATED WITH THEIR OWN MATURE HEIGHTS AND WEIGHTS

Age	Height				Weight			
	Boys		Girls		Boys		Girls	
	N	r	N	r	N	r	N	r
Birth	—	—	—	—	23	.39	23	.51
Mo. 1	18	.64	20	.37	19	.19	20	.31
3	23	.53	23	.51	23	.18	23	.51
6	22	.72	21	.58	23	.32	21	.40
9	21	.59	23	.58	21	.12	23	.46
Yr. 1	21	.72	19	.65	23	.26	22	.51
2	21	.68	19	.68	21	.42	21	.62
3	21	.83	19	.85	23	.44	22	.59
4	21	.85	19	.82	20	.32	23	.63
5	21	.86	19	.84	22	.44	23	.68
6	21	.88	19	.85	22	.42	24	.79
7	21	.87	19	.85	23	.52	23	.81
8	21	.87	19	.84	22	.42	22	.87
9	21	.89	19	.83	23	.52	22	.86
10	21	.89	19	.80	22	.44	24	.87
11	21	.88	19	.64	23	.60	21	.81
12	21	.88	19	.77	20	.58	23	.78
13	21	.82	19	.88	22	.67	19	.53
14	21	.78	19	.73	19	.54	20	.86
15	21	.74	19	.98	20	.53	20	.92
16	21	.79	19	.997	22	.65	22	.95
17	21	.92	19	.999	20	.76	23	.94
18	21	.98	—	—	21	.89	—	—

TABLE XI.—CORRELATIONS FOR CONSISTENCY OF TWO BODY BUILD INDICES: EIGHTEEN YEARS COMPARED WITH THREE EARLIER AGES

Age in Months	Stem Length/Height				Weight/Height^2			
	Boys		Girls		Boys		Girls	
	N	r	N	r	N	r	N	r
3	21	.47	17	.43	21	.10	18	.22
12	21	.08	19	.20	21	−.09	20	.81
36	21	.45	19	.61	21	.08	19	.71



FIGS. 3 and 4. The Children's Sigma Scores relative to the Means for the Berkeley Growth Study for Successive Ages are shown by Dots and a Smoothed Curve. For Comparison the Parents' Sigma Scores relative to the Parents in this Group are indicated by a Straight Horizontal Line.

mean) is indicated by a broken horizontal line. As would be expected from the correlations, most (though not all) children's scores tend to settle at a level near that of their parents. But the individual curves bring out another fact, that some children reach that level much sooner than others.

The different patterns in physical growth can be shown to be due, in part, to differences in rates of maturing. For example, case 18F matured very slowly: her skeletal age was retarded for all ages after two years, usually lagging behind at least a full year. When she finally reached maturity she had attained a tall stature similar to that of her parents. Case 8M was rather consistently short, like his parents: his skeletal age was average until thirteen years, after which it lagged somewhat, apparently postponing his reaching an eventual status slightly taller than his parents. Case 3M, on the other hand, was accelerated at least a year at all ages after two years, and he tended to be relatively taller than his parents until he approached skeletal maturity. After sixteen years his scores fall below the parental line.

We do not understand clearly the nature of the maturing processes in mental functions. There could be inherent differences in rates, and also differential environmental factors operating to retard or enhance intellectual growth. But if we think of these children's scores at twenty-one years as measures of their mature intellectual status, it is clear that they follow varied paths in attaining that status. Case 8M had an early short period of rapid development, and after two years maintained a position that is close to both his parents' and his own mature level. Case 5F started out slowly, but caught up to her own adult and her parental level by three and one-half years. Case 14F, however, after a very precocious start settled more gradually to a position which never dropped quite as low as her parents' educational status, but stabilized at about eight years.

Examination of all the individual curves (in addition to those shown here) reveals a great variety of trends, with examples of (relatively) rapid or slow growth occurring at all ages, both in heights and in intelligence scores.

SUMMARY AND INTERPRETATION

Most of the physical comparisons show increasing parent-child similarities as the children grow older. However, there appear to be

some differences, both in the amounts of relationship and in the age-trends. Because the samples are small, it is possible that these differences are peculiar to the samples. If the samples are representative in these respects, it would appear that the within-sex relationships are stronger than those between parent and child of the opposite sex. Also, the mother-daughter relations seem to be closer than the father-son relationships.

The trends in mental relations differ in several ways from those in physical relations. The children themselves are more stable in the physical measurements. But the trends of increasing similarity to their parents seem more consistent in the mental scores. Althought there are no sex differences in the parent-child *r*'s for the mental factors, there appear to be some marked sex differences in the physical factors. The analogy between the mental and physical relations is, obviously, not a close one. Furthermore, a variety of questions could be asked about genetic relations in physical growth which cannot be resolved here.

From these data it does, however, seem reasonable to assume the existence of a hereditary core of parent-child similarities in both mental and physical characters, even though such similarities may not be evidenced during the first year or so of the child's life. The similarities tend to increase as the children approach the mature status at which the characteristics were observed in their parents. Inherent differences in rates of development during infancy could account, at least in part, for the early lack of parent-child similarity.

Other factors operative in the reduced *r*'s at early ages are clearly observable in the phenomena of physical growth. These are the differences in the infants' body proportions from those of adults, and the differential growth rates of different parts of the body. For example, an infant's total height is made up in large part by his stem length, while the adult's height is determined to a much greater extent by the length of his legs. In the Berkeley Growth Study, the mean stem-length/height ratio drops from .65 at three months to .51 between twelve and fourteen years (signifying relatively longer legs), and then rises slightly to about .53 in the young adult. The transverse dimensions of the body, as contrasted with the longitudinal dimensions, grow more rapidly during the first year after birth and again during early adolescence.

Analogous age changes in organization could be expected in any

complex function during the process of its development (*4*). If we accept this position, then we may assume that differences in the mental organization of the infant and the adult could contribute to the changing parent-child relations in mental ability. Credence is lent to this hypothesis by the study of Skodak and Skeels (*15*) who report increasing mother-child *r*'s which cannot be attributed to environment. The effects of environmental influences should, consequently, be evaluated in the light of both age changes in organization (or relative maturity of the different parts or factors) and individual differences in over-all rates of maturing.

Although the sample used here is too small to allow for anything but tentative interpretations, it seems very likely that the changing parent-child relationships exhibited here are widespread. They deserve more extensive investigation in other populations and in other characteristics.

BIBLIOGRAPHY

1) N. Bayley. "*Studies in the Development of Young Children.*" Berkeley: University of California Press, 1940, pp. 45

2) ———— "Factors influencing the growth of intelligence in young children." *Thirty-Ninth Yearbook National Society for the Study of Education*, 1940, 39, 49–79.

3) ———— "Consistency and variability in the growth of intelligence from birth to eighteen years." *J. Genetic Psychology*, 1949, 75.

4) ———— "Development and maturation." *Theoretical Foundations of Psychology*, (H. Helson, Editor) New York: Van Nostrand Co. 1951, 145–199.

5) N. Bayley and F. C. Davis. "Growth changes in bodily size and proportions during the first three years; a developmental study of sixty-one children by repeated measurements." *Biometrika*, 1935, 27, 26–87.

6) N. Bayley and H. E. Jones. "Environmental correlates of mental and motor development: a cumulative study from infancy to six years." *Child Develop.*, 1937, 8, 329–341.

7) H. S. Conrad. "On kin resemblances in physique vs. intelligence." *J. Educ. Psychol.*, 1931, 22, 376–382.

8) H. S. Conrad and H. E. Jones. "A second study of familial resemblance in intelligence: environmental and genetic implications of parent-child and sibling correlations in the total sample." *Thirty-Ninth Yearbook of the National Society for the Study of Education, Part II*, 1940, 97–141.

9) H. Gray. "Prediction of adult stature." *Child Develop.*, 1948, 19, 167–175.

10) M. P. Honzik. "Age changes in relationship between certain environmental variables and children's intelligence." *Thirty-ninth Yearbook National Society for the Study of Education, Part II*, 1940, 185–205.

11) H. E. Jones. "A first study of parent-child resemblance in intelli-

gence." *Twenty-Seventh Yearbook, National Society for the Study of Education*, 1928, Part, I61–72.

12) H. E. Jones and H. S. Conrad. "The growth and decline of intelligence: a study of a homogeneous group between the ages of ten and sixty." *Genet. Psychol. Monog.*, 1933, 8, 233–297.

13) ———— *"Parent-Child Resemblance in Mental Abilities as a Function of Age."* Unpublished manuscript at the Institute of Child Welfare, University of California.

14) Q. McNemar. *"Psychological Statistics."* New York: John Wiley & Sons, 1949, pp. 364.

15) M. Skodak and H. M. Skeels. "A final follow-up study of one hundred adopted children." *J. Genet. Psychol.*, 1949, 75, 85–125.

TWO METHODS OF TEACHING ENGLISH COMPOSITION TO RETARDED COLLEGE FRESHMEN[1]

RAY CHARLES MAIZE

Air University, Maxwell AF Base, Alabama

This study was established to provide a comparison between a 'language by experience' procedure and a 'grammar drill' method of teaching composition to retarded students in English A sections at Purdue University. Its purposes were: (1) to determine which method teaches the student more effective habits of language application as measured by objective testing and sample writing; (2) which method requires more extensive outside work and preparation on the part of the instructor; and (3) what relative factors, such as reading skill and mental levels, influence the student's ability to achieve success.

Sections of English A at Purdue University were used for the experiment. These slow and retarded students were separated from the other freshmen at Purdue by the scores they received on the American Council on Education Psychological Examination, the Purdue University Placement Test in English, and an original theme written during the entrance examination period. If their percentiles on the ACE Psychological Examination and the Purdue Placement Test in English averaged below 30 and their grade on the original theme (graded by the English faculty) was below 3 (passing), then they were assigned to English A, a non-credit, one-semester course which they must take before going on to English 1.

Recognizing the deficiencies of these students, the English faculty at Purdue had established English A as a remedial measure to bring the student to some degree of proficiency which would enable him to pass English 1, the required course for freshman credit in English. Taught according to the program established by the English department, English A consisted of workbook drill of grammar, punctuation, and spelling; and a weekly theme, corrected by the instructor. These procedures were supposed to be specifically de-

[1] This article is based on the writer's doctoral dissertation, Purdue University, 1952. Grateful acknowledgment is made to Professor Einar R. Ryden under whose direction this work was done.

signed to change the language habit patterns of slow, retarded freshmen.

To find out how well English A achieved the purpose for which it was established, permission was granted by the head of the English department to conduct an experiment in which the officially sanctioned method of instruction would be contrasted against a writing laboratory method. Accordingly, one group, the control, was organized and taught using the method established by the English department. The other group, the experimental, was taught using the writing laboratory method.

Inherent in this latter method would be the need to free the instructor from extensive theme reading, for it was recognized at the beginning that merely doubling or tripling the number of student mansucripts would provide no answer for the one great burden of teachers of English: how to get student writing activity without increasing the already too heavy load of the teacher. The second objective, then, of the experiment would be the need to prove that the teacher does not have to edit all student writing, that more effective results can be obtained if the teacher deliberately refrains from overdoing such supervision while at the same time encouraging voluminous efforts by the students.

The third objective of the experiment lay in the need to discover the special problems of the retarded students. What factors, in other words, had greatest relevance to their success in the course? Was it reading skill, mental level, vocabulary, attitude, or feeling of failure or success? Just as important, therefore, as the degree of control over his language habits which the experiment should reveal was the necessity to learn more about the individual who was to be taught.

STRUCTURE OF THE EXPERIMENT

From the freshmen students assigned to English A sections at Purdue during the year 1949–1950, one hundred and forty-nine were selected at random to become a part of this experiment to determine relative merits of two different approaches to the teaching of language skill on a college freshman level.

The method of selection for membership in any one of the groups has already been described. On the basis of these factors, students were registered by chance in the sections used for the experiment: seventy-four to the control group and seventy-five to the experi-

mental group. The control group was taught by the grammar workbook drill method; the experimental group was taught by the writing laboratory method.

During the semester of instruction, the following tests were administered: the Terman-McNemar Test of Mental Ability; the Henmon-Nelson Test of Mental Ability; the Iowa Silent Reading Test; the Rinsland-Beck Natural Test of English Usage, Tests I, II, and III; the Seashore Recognition Vocabulary Test; essays on an identical subject written before and after the period of instruction; the Horrall College Experience Scale; the Silance-Remmers Scale to Measure Attitude Toward Subject; and the Silance-Remmers Scale to Measure Attitude Toward Teacher.

To determine the amount of change and the significance of change before and after instruction, six units of measurement were established. These units were: (1) identical forms of the Seashore Recognition Vocabulary Test to determine change in total vocabulary; (2) equivalent forms A and B, Test I of the Rinsland-Beck Natural Test of English Usage, to measure change in ability to capitalize and punctuate; (3) equivalent forms A and B, Test II of the Rinsland-Beck to determine change in ability to work with diction, tense, pronoun reference, and verb forms; (4) equivalent forms A and B, Test III of the Rinsland-Beck to measure change in ability to improve sentence structure, organize logically, and write effectively; (5) combined scores for Test I, II, and III of the Rinsland-Beck to obtain composite language results; and (6) average of numerical theme grades awarded by two different English instructors after reading student manuscripts on an identical subject assigned before and after instruction.

Since one of the stated purposes of this study was to determine what relative factors have influence on the student's ability to achieve success, two different intelligence tests were used, in addition to the ACE Psychological Examination. The Iowa Silent Reading Test was also used to discover the relationship of scores in reading to intelligence scores and to other evaluation forms.

It was also considered important to find out what effect, if any, the personality and enthusiasm of the various instructors might have on the results. In addition, the attitude of depression or buoyancy, caused by success or failure, might also have an effect on the findings. For these reasons the Silance-Remmers Scale to Measure Attitude Toward Subject, the Scale to Measure Attitude

Toward Teacher, and the Horrall College Experience Scale were used.

During the period of instruction the control group wrote fourteen themes of approximately two hundred and fifty words in length. These themes were read and corrected by the instructor outside the classroom. The experimental group wrote forty themes of approximately two hundred and fifty words in length. The instructor did not read or correct any of this writing outside the regularly scheduled class hours; instead analysis and comment were provided in the writing laboratory by fellow students and the instructor.

ANALYSIS AND INTERPRETATION OF DATA

In this research the control and experimental groups were considered equal in ability at the beginning of the study. Data for indications of equality were obtained from the mean of the raw scores received by the students on the ACE Psychological Examination, the Purdue Placement Test in English, and the numerical theme grade. The t-ratios which were obtained for the difference between these means were not statistically significant according to t-tables.

MEASURE OF CHANGE

Table 1 indicates the results obtained from the six units of measurement which were established to determine the amount of change and the significance of change for the control and experimental groups.

As shown in Table 1, changes on the vocabulary test were favorable and significant at the one per cent level of confidence for both groups. A t-ratio for the difference between the changes, however, was not statistically significant. The difference between changes of the two groups could, therefore, be due to chance.

On Test I of the Rinsland-Beck the change was significant at the one per cent level of confidence for both groups. A t-ratio for the difference between changes, however, was significant at the five per cent level of confidence. The change for the experimental group, therefore, was significantly larger.

On Test II of the Rinsland-Beck the change for the control group was significant at the 5 per cent level of confidence, for the experimental group at the one per cent level of confidence. A t-ratio for the difference between changes was acceptable at the five per cent

Table I. Means and t-Ratios for Control and Experimental Groups on Measurements to Determine Change before and after Instruction

	Mean of Control		Mean of Experimental		t-ratio	
	Before	After	Before	After	Contr	Exp
Seashore Recognition Vocabulary Test	61.1	58.6	62.5	57.8	−3.6‡	−5.9‡
Test I, Rinsland-Beck Natural Test of English Usage	55.4	59.0	55.0	62.8	5.1‡	9.8‡
Test II, Rinsland-Beck Natural Test of English Usage	53.1	55.2	49.3	54.9	2.3†	5.6‡
Test III, Rinsland-Beck Natural Test of English Usage	29.5	28.3	24.9	33.2	−1.2	5.9‡
Combined Scores, Rinsland-Beck Natural Test of English Usage	138.1	142.4	129.2	150.9	3.1	12.8‡
Numerical Grades Given By Theme Readers	9.7	9.1	10.1	11.9	−1.2	6.0‡

† 1% level of confidence.
‡ 5% level of confidence.

level of confidence. The experimental group, therefore, achieved greater change than the control group.

On Test III of the Rinsland-Beck the control group showed no change that was statistically significant. The experimental group, on the other hand, showed change which was significant at the one per cent level of confidence. The t-ratio for the difference of change was also significant at the one per cent level. On this form, directly testing the student's ability to write, the evidence was all in favor of the experimental group.

On the Combined Scores of the Rinsland-Beck the total change for both groups was statistically significant at the one per cent level of confidence. The t-ratio for the difference of change, however, was significant at the one per cent level of confidence in favor of the experimental group.

On Numerical Grades Given by Theme Readers, the control group showed no significant change. The experimental group, on the other hand, did show positive gain in ability to write, significant at the one per cent level. The t-ratio for the difference of changes was also significant at the one per cent level.

CORRELATION OF FACTORS

One of the stated purposes of this study was to determine the relationship to one another of the many factors involved, in an effort to understand more clearly the nature of the teaching task and to obtain the necessary insight into those factors which seemed to contribute most to success or failure on the part of the student.

The following details are worthy of note:

a) The mean chronological age of the control group was 19.5 years; the mean age of the experimental group was 19.6 years.

b) The mean raw scores for both control and experimental groups on the Henmon-Nelson Test of Mental Ability were reported by the authors of the test at the 40th percentile, indicating that the students who form the population for this study were at least in the lower half of any population reported on by standardized tests.

c) When the mean raw scores on the Terman-McNemar Test of Mental Ability are compared with tables provided by the authors, the results show a mean IQ for both groups of 99 and a percentile ranking of 48.

d) The highest degree of correlation found between the Iowa Silent Reading rate and any other evaluation factor was that of .51 with the Q score on the ACE Psychological Examination. Otherwise the rate of reading seemed to bear little relationship with anything else.

e) The mean of the standard comprehension score as reported for the control and experimental groups on the Iowa Silent Reading Tests show a percentile ranking of 16 and a grade equivalent of 10.2, indicating a two to three year retardation in reading ability for students who are supposedly working on a 13th grade level.

EFFECT OF ATTITUDES

It was considered important to the results of this study to find out what effect, if any, the personality and enthusiasm of the various instructors might have. In addition, the attitudes of depression or buoyancy, caused by success or failure in the course, might also

have an effect upon the findings. It therefore became necessary to discover whether such moods differed between the control and experimental groups. On the Silance-Remmers Scale to Measure Attitude Toward Subject and Attitude Toward Teacher no difference could be found between attitude reactions of students in control and experimental groups. Similarly, no difference could be found on the Horrall College Experience Scale to measure depression or buoyancy. The conclusion was drawn that the lack of significant difference in measurement of attitude or mood indicated a lack of influence on the evaluation techniques used in this study.

SUMMARY AND CONCLUSIONS

As a result of the methods of instruction and the testing conducted before, during, and after the period of instruction, these conclusions can be drawn:

a) The control group showed statistically significant change measured before and after instruction on vocabulary, mechanics, and grammar. The experimental group, on the other hand, showed statistically significant change as measured before and after instruction on vocabulary, mechanics, grammar, controlled composition, and free composition.

b) Furthermore, the experimental group showed a statistically significant larger change on all measurements except vocabulary. All other measurements used in this study, therefore, showed clear evidence of superiority on the part of the experimental groups as the result of the methods of instruction.

c) All data revealed that students were of limited mental capacity and possessed an initial low level of ability in language use, vocabulary, and reading. In addition, changes for both low-level groups were slight, even after a semester of instruction.

d) No evidence could be found of the effect of student attitude or mood on the results of this study.

DIFFERENTIAL RESPONSES AMONG COLLEGE STUDENTS ON THE VOCABULARY SUBTEST OF THE WECHSLER INTELLIGENCE SCALE

CHALMERS L. STACEY and S. WILLIAM SPANIER

Syracuse University

Gerstein (*1*) in an article in the *Journal of Consulting Psychology* in 1949 suggests a relationship between the use of language and the underlying cognitive process involved. She proposes placing more emphasis on this relationship instead of merely determining whether or not the subject has given the correct definition of a word on a vocabulary test. That is, one should, in addition to determining whether the definition is correct or incorrect, attempt to differentiate the type of definition which is considered correct.

Reichard, Schneider and Rappaport (*2*) using the Goldstein-Weigl (*3*) color-sorting test as an instrument for investigating concept formation in children arrived at the conclusion that there exist three basic methods for forming concepts: (a) the concretistic, (b) the functional and (c) the conceptual.

Gerstein on the basis of the Reichard, Schneider and Rappaport study feels that the same three methods of concept formation which were exhibited in that study to solve a performance problem may also be used in classifying the definitional levels of responses to verbal stimuli. She suggests the following three classifications which parallel the aforementioned three: (a) the concretistic or descriptive method which is considered to be at the lowest level of verbalized definitions (the most primitive level), e.g., "an apple is red" or "a diamond is very hard and glitters, shines and sparkles"; (b) the functional or usage method, a more complex level and considered to be at the next higher level of verbalized definitions, e.g., "a coat is something you wear to keep warm" or "a microscope is used to make observations of minute objects"; and (c) the categorical or conceptual or abstract method, the most complex method and considered to be at the highest level of verbalized definitions, e.g., "wood is a fuel" or "a bicycle is a means of transportation".

STUDIES MADE ATTACKING THIS PROBLEM

In two separate studies, Stacey and Portnoy (*4*, *5*) attempted to test the hypothesis put forth by Gerstein. Both studies used groups of individuals that were subnormal in intelligence as determined by the Wechsler Intelligence Scale. In their first study, when comparing the results of two groups of children, one classified as borderline and the other as mentally defective, they arrived at the following results, having used the Vocabulary Subtest of the Wechsler Intelligence Scale for Children for the verbal stimuli: (1) the borderline group (higher on the scale of intelligence) gave a greater proportion of descriptive type definitions than did the mental defective group—in fact, the two groups were significantly different at the one per cent level, (2) the mental defective group gave a greater proportion of the functional type of definition—this difference was also significant at the one per cent level, and (3) there was no significant difference between the two groups where the conceptual type of definition was concerned though the borderline group did show a slightly higher proportion of this type of definition.

In their second study, using three groups of adults, one classified as dull-normal, one as borderline and the last as moron, they found no significant differences in terms of the critical ratios between any two groups of the three in any of the three categories (descriptive, functional or conceptual). Nevertheless, the same general trend was evident. As in the first study, the higher level of intelligence gave a greater proportion of descriptive type definitions than did the next lower level of intelligence, and a lower proportion of functional type definitions than did the next lower level of intelligence, and a slightly higher proportion of conceptual type of definitions than did the next lower level of intelligence. Stacey and Portnoy, in this same study, then compared the two groups which comprised the two studies but omitting the dull-normal group of the second study, i.e., they compared the adult moron (of the second study) with the child moron (of the first study) and the adult borderline (of the second study) with the child borderline (of the first study) thereby holding intelligence constant and age as a variable. The following results were obtained: (a) the older age groups gave a greater proportion of descriptive and abstract types of definition, the difference being significant at the one per cent level, (b) the younger age groups gave a greater

proportion of functional definitions. Again the difference was significant at the one per cent level.

The results of these two studies appear directly to contradict Gerstein's hypothesis of the levels of methods of concept formation with the exception of the abstract or conceptual method. Gerstein's hypothesis is thus contradicted in terms of levels of intelligence and the Reichard, Schneider and Rappaport hypothesis (which includes the statement that children think at the concrete or descriptive level up to about age eight, and somewhere between the ages of eight and nine the functional level is reached, and at about the age of eleven the conceptual or abstract level is reached) is thus contradicted in terms of age.

In another study, Stacey and Markin (*6*) used the Similarities Subtest of the Wechsler Bellevue Intelligence Scale Form I with three groups of adults classified as (1) mental defective, (2) borderline and (3) dull-normal. Again as in the two previously mentioned studies by Stacey and Portnoy, the same trend may be noted. The greater proportions of descriptive responses were made by the higher levels of intelligence—the difference between the mental defectives and the dull-normal groups being significant at the one per cent level; the difference between the other groups, though not significant, showed the same trend. Also, the greater proportion of functional responses were made by the lower levels of intelligence —the difference between the mental defectives and dull-normal being significant at the one per cent level; the difference between the mental defectives and the borderline groups being significant at the five per cent level and the difference between the borderline and dull-normal groups, though not significant, again showed the same trend. The differences between the three groups at the abstract level were much too small to draw any conclusions.

Since all the results thus far have been gained through the use of individuals of subnormal intelligence as subjects, it is the purpose of this paper to further test Gerstein's hypothesis, but this time using as subjects one hundred eighty-eight college juniors and seniors.

METHOD

The vocabulary section of the Wechsler Bellevue Intelligence Scale was administered to one hundred eighty-eight college students. The responses given by the students to each of the forty-

five words in this section were recorded verbatim. The response sheet of each student was graded in accordance with the instructions cited in the Wechsler Bellevue Manual. Two separate groups of students were established by selection of the highest twenty-five per cent of the grades and the lowest twenty-five per cent of the grades (forty-seven students in each group). Each correct definition given by each student for each of the nouns (only nouns were used for analysis) listed on the response sheet was then broken down into one of the three categories: descriptive, functional or conceptual.

Each breakdown was then recorded in its proper column on a chart, and the total number of correct responses for each student also recorded on the chart. A grand total of correct responses was then calculated and recorded on the same chart. Separate totals were also calculated for each of the three categories and the proportions of each of the three separate categories to the grand total number of correct responses computed. This procedure was followed separately for the two groups—the upper twenty-five per cent and the lower twenty-five per cent. It might also be said that a third group was established from the middle twenty-five per cent of the grades and the same procedure followed with this group. Of course, any significant difference that might exist would come to light in the critical ratios computed between the high and low groups. The middle group should only serve to bridge the gap between these other two groups, and it is for this reason only that the third group was established. Critical ratios were then computed between the proportion of responses given by the three groups of students in each category, i.e., a critical ratio was computed for the proportion of descriptive responses in the high group as compared to the proportion of descriptive responses in the low group, and also the proportion of descriptive responses in the middle group was compared to the proportion of descriptive responses of both the high and low groups separately. The results of these comparisons are presented in Table I.

RESULTS

1) The proportion of descriptive responses made by the top and bottom groups differed significantly at the one per cent level. There was no significant difference between the proportions of descriptive responses given by the top and middle groups, but

TABLE I.—FREQUENCY AND PROPORTIONS OF LEVELS OF RESPONSE TO THE VOCABULARY SECTION OF THE WECHSLER-BELLEVUE INTELLIGENCE SCALE

Category	Top 25%		CR	Middle 25%		CR	Bottom 25%		CR	Top 25%	
	Freq.	Prop.		Freq.	Prop.		Freq.	Prop.		Freq.	Prop.
Abstract	1043	.74	4.05*	880	.67	2.11†	741	.63	6.11*	1043	.74
Functional	23	.02	6.66*	90	.07		78	.07	6.49*	23	.02
Descriptive	339	.24	1.21	347	.26	2.22†	347	.30	3.43*	339	.24
Total	1405			1317			1166			1405	

* Significant at the one percent level.
† Significant at the five percent level.

the trend was still evident—an increase in the proportion of descriptive responses given by the higher group to the next lower group. The proportion of descriptive responses between the middle and bottom groups differed significantly at the five per cent level.

2) The proportion of functional responses made by the top and bottom groups differed significantly at the one per cent level. The proportion of functional responses between the top and middle groups differed significantly at the one per cent level, but there was no difference at all in the proportions of functional responses between the middle and bottom groups.

3) The proportion of conceptual responses made by the top and bottom groups differed significantly at the one per cent level. The proportion of conceptual responses between the top and middle groups differed significantly at the one per cent level and the proportion of conceptual responses between the middle and bottom groups differed significantly at the five per cent level.

4) Each of the groups gave the greatest proportion of responses in the conceptual category, the next highest proportion of responses in the descriptive category and the lowest proportion in the functional category.

SUMMARY

The responses of the three groups of college students (groups arbitrarily established on the basis of scores on the Vocabulary Subtest of the Wechsler Bellevue Intelligence Scale) were analyzed as to the proportions of their vocabulary responses that fell into the descriptive, functional or conceptual categories.

According to the Gerstein hypothesis the descriptive type of definition is the most primitive, the functional type of definition is the next highest level, and the conceptual type of definition is the most complex. We should therefore expect that in any study of this kind we would find that the groups of higher intelligence levels would give (1) a greater proportion of conceptual type definition than would the groups of lower levels of intelligence, (2) a somewhat higher proportion of functional type definition than would the groups of lower levels of intelligence and (3) a lower proportion of the descriptive type of definition than would the groups of lower levels of intelligence. But in the three studies mentioned earlier in this paper, the results substantiated Gerstein's hypothesis in only one of the three categories—the conceptual. The other two categories—the functional and the descriptive —ran directly counter to the hypothesis set forth by Gerstein.

The results of this study are in agreement with the results of the previous studies on only two of the three categories—the conceptual and the functional. The discrepancy lies in the descriptive category. We therefore find that the results of this study substantiate Gerstein's hypothesis on two of the three categories—the conceptual and the descriptive. That is to say that: (1) the top group gave a higher proportion of conceptual responses than did the middle group and the middle group a higher proportion than did the bottom group, (2) the top group gave a lower proportion of descriptive responses than did the middle group and the middle group a lower proportion than did the bottom group; in contrast to Gerstein's hypothesis the top group gave a lower proportion of functional responses than did the middle group and there was no difference at all in the proportion of functional responses between the middle and the bottom group.

The discrepancy which exists between these results and those of the three previous studies—namely, the discrepancy in the descriptive category—may, of course, be due to the fact that in the previous studies subjects of subnormal intelligence were used whereas the sample in this study was composed of college students. Certainly a great deal more research is necessary before any absolute conclusions may be drawn or even assumed. However, the results of all four studies do leave in doubt that part of Gerstein's hypothesis which states that the functional type of definition is of a higher level than is the descriptive. But here again

further research is necessary before these results may be considered conclusive.

REFERENCES

1) R. A. Gerstein. "A suggested methld for analyzing and extending the uses of Bellevue-Wechsler vocabulary responses." *Journal of Consulting Psychology*, 1949, 13, 366–370.

2) S. Reichard, M. Schneider, and D. Rappaport. The development of concept formation in children." *American Journal of Orthopsychiatry*, 1944, 156–161.

3) K. Goldstein and M. Scherer. "Abstract and concrete behavior: An experimental study with special tests." *Psychological Monograph*, 1941, No. 2.

4) C. L. Stacey and B. Portnoy. "A study of the differential responses on the vocabulary subtest of the Wechsler Intelligence Scale for Children." *Journal of Clinical Psychology*, 1950, 6, 401–403.

5) ———— "A study of the differential responses on the Vocabulary Subtest of the Wechsler Intelligence Scale." *Journal of Clinical Psychology*, 1951, VII, No. 2, 144–148.

6) C. L. Stacey and K. Markin. "A study of the differential responses among three groups of subnormals on the similarities subtest of the Wechsler Intelligence Scale." *American Journal of Mental Deficiency*, 56, No. 2, Oct. 1951.

ITEM VALIDITY AND RESPONCE CHANGE UNDER TWO DIFFERENT TESTING CONDITIONS[1]

DOUGLAS G. SCHULTZ

Educational Testing Service

Princeton, N. J.

It has often been demonstrated that various responses to the usual self-description inventories and questionnaires can be elicited by altering the motivational pressures under which the instrument is administered. Bennett and Gordon (*1*), Ruch (*5*), and Schott (*6*) have all reported studies in which persons who took this type of test as part of the selection process for school or job gave considerably more favorable responses than persons who took the test under non-selection circumstances. Using the same questionnaire with which this report will be concerned, Myers (*3*) found a large number of statistically significant differences between the mean item responses of students accepted for college and those of college applicants. He also found that the applicants presented themselves generally as being more sober and industrious and better adjusted than the accepted students. Results obtained by Hollingworth (*2*), Simpson (*8*), and Van Wagenen (*9*) also support the point. When the answers given to inventories and questionnaires may have a decisive bearing on the attainment of some desired goal, it is apparent that many people either intentionally or unintentionally tend to respond in a manner which they think or feel will present the best opportunity for achieving that goal.

The study reported in this paper examined such shifts in response from the standpoint of the individual item. (Throughout the paper, responses to items and validities of items will be discussed.) The purpose of the research was to determine how changes in the item responses to an attitude-interest questionnaire administered under selection and non-selection testing conditions are related to the validities of the questionnaire items for an academic criterion. Specifically the two problems investigated could be

[1] This paper was read at the American Psychological Association meeting in Washington, D. C., on September 2, 1952.

phrased as follows:

1) Does the average response to an item change more (from non-selection to selection administrations), if the item is of high validity than if it is of low validity?

2) Do the items to which the average response is changed most also decrease most in validity?

HYPOTHESES

With regard to the first of these problems, it was anticipated that the correlation between mean item response change and external item validity would be zero, on the grounds that students seeking admission to college would exhibit no consistent insight into the way the items operate. It was felt that, in their general desire to give the most favorable impression, they would not differentiate between the more and less valid items and that therefore the mean change would be similar for items of all levels of validity.

Turning to the second issue, widespread response shifts from non-selection to selection groups were expected on all items and it was thought that these shifts might raise or lower item validities by chance. But it was hypothesized that there would be some tendency for the larger changes in item response in either direction to be associated with the greater reductions in item validities. Generally, then, change in mean response would be related to change in item validity only to a small extent.

SUBJECTS

Members of two successive freshman classes at a women's liberal arts college of an eastern state university were the subjects for the study. There were three hundred fifty-five women in the class of 1951 and three hundred forty-five in the class of 1952.

PROCEDURE

For several years the Educational Testing Service has been attempting to develop a questionnaire which would measure non-intellectual factors associated with academic achievement in college.[2] The study, supported by the College Entrance Examination

[2] The development and validation of this questionnaire over a three-year period are reported elsewhere in two papers (*4*, *7*).

Board, has been directed primarily at evaluating characteristics of women entering college. A questionnaire of three hundred twenty-eight items was constructed and administered to the three hundred fifty-five members of the freshman class, the Class of 1951, as a group three weeks after their arrival at college. One hundred and forty-five of the same items were arbitrarily selected for reproduction in a booklet which was administered individually to applicants for the next year's freshman class, the Class of 1952, as a part of the normal application for admission; three hundred forty-five women were admitted to this class. Thus, members of the first class responded after they had been safely admitted to the college, whereas respondents in the second group were seeking admission to college. These are the postadmission and preadmission groups, respectively.

The one hundred and forty-five items in the questionnaire were of the usual inventory type. For example, one item asked: "Do you feel that the higher the grades a girl gets in college the more she will amount to after college?" The five possible responses were: (1) Yes, (2) Probably Yes, (3) Uncertain, (4) Probably No, and (5) No. In other items the respondent was to indicate how appealing certain topics of conversation might be to her. In some items she was to give the degree of importance to her of various reasons for attending college. Over all, the content covered the student's intellectual interests, motivation for attending college, and teacher relations and study habits.

Each of the items had four, five, or seven possible responses, numbered from 1 to 4, 1 to 5, or 1 to 7. Since this numerical order corresponded to a logical order of responses, a linear response continuum was assumed and the score of a person on an item was taken to be the number of the response which was checked. Thus, a respondent checking 'Probably No' in reply to the question cited above would receive a score of 4 on the item.

Using these scores, the average response for each of the one hundred and forty-five items in each of the two classes was computed. The Class of 1951 (postadmission) item mean subtracted from the Class of 1952 (preadmission) item mean was taken as an index of the change induced in the response to that item by the different conditions of administration. The difference for any item could be either positive or negative, depending only on whether

the average of the response numbers checked by the Class of 1952 was higher or lower on the arbitrary response scale than the average of the numbers checked by the Class of 1951. It should be noted that the same individuals are not being compared under two testing conditions here nor are two groups of matched pairs of students. Assuming that the two classes were selected in the same manner from the same general population, however, it is reasonable to compare the mean score of an item for one class with the mean score for the other class for the purposes of the study.

The criterion used for validating the items was the so-called 'achievement index', which was derived for each individual by taking the difference between her actual Freshman-Year College Grade-Point Average and her most probable Grade-Point Average predicted from the two scores of the College Board Scholastic Aptitude Test. (The multiple correlations of the Scholastic Aptitude Test verbal and mathematical sections with Freshman-Year Grade-Point Average were .51 in the Class of 1951 and .52 in the Class of 1952.) The product-moment correlation of each of the one hundred and forty-five items with the criterion was obtained. It was possible to calculate two such validity indices for each item, one in each class. It should be noted that since the items were of the common inventory type there was no 'right answer' in the sense of factual items and so there was no expected sign for any validity coefficient; either negative or positive values might be equally satisfactory and both were found in the present case.

Using the mean response differences and item validities described, two kinds of analyses were then carried out:

(a) Each item validity index in the Class of 1951 (non-selection) was plotted against the difference between the two class mean responses for that item; a similar plot was made with the Class of 1952 (selection) validity indices.

(b) The change in item validity from one class to the other was plotted against the mean item response difference. In considering the change in item validity a rise in the direction of the non-selection group's value (Class of 1951) was taken as a positive change or 'rise' in validity, a drop toward or beyond zero in the other direction as a negative change or 'drop'.

Since items of four, five, and seven choices were included, plots were also made substituting the critical ratio of the difference for

the mean item response difference. The item validity r's were not transformed into z's because all of them were low and so in the range where the transformation has little effect.

RESULTS

For the results, the two kinds of analyses will be considered separately. First, on the problem of item validity and item response change, the correlations computed from the scatterplots were as follows:[3]

In the Class of 1951 (Postadmission)

Item validity vs. mean item response change: +.13

Item validity vs. CR of mean item response change: +.13

In the Class of 1952 (Preadmission)

Item validity vs. mean item response change: +.21

Item validity vs. CR of mean item response change: +.24

The positive sign means that the largest item response changes in the direction indicated as desirable by the sign of the item validity coefficient are associated with the items of highest validity. Assuming that all the conditions underlying use of the common formula for the standard error of a zero correlation are fulfilled here, the N being one hundred and forty-five items, the two values of +.13 do not differ significantly from zero in the statistical sense, the +.21 almost reaches the .01 level of significance, and the +.24 does reach the .01 level. Although all four coefficients cited are positive, the pertinence of this agreement from the standpoint of statistical significance is somewhat reduced by the fact that the coefficients are by no means independent of one another.

As a matter of interest, the correlations using the critical ratio in an attempt to adjust for the variations in number of possible

---

[3] The means and standard deviations for the four variables involved were:

	M	SD
Item validities (Class of 1951)	−.02	.06
Item validities (Class of 1952)	.00	.06
Differences between mean item responses	+0.05	0.35
CR's of differences between mean item responses	+0.57	3.99

A note should perhaps be added to clarify these item validities statistics. The item validities were distributed approximately symmetrically around zero, both positive and negative values being found as mentioned above, so the means of the item validities were close to zero as reported here. The most valid items, of course, would be those with the highest validity coefficients regardless of sign.

responses do not differ essentially from those using the straightforward mean response difference. Also, the differences between comparable coefficients from the two classes are neither large nor statistically significant.

When the scatterplots for the four correlations were examined, the picture seemed to be a very common one with no suggestion of a curvilinear trend of any kind.

Turning to the second problem, the correlation between drop in item validity from the post- to preadmission class and change in mean item response was +.01. Substituting the critical ratio for the mean response difference resulted in a correlation of −.02. Thus, no association was found between decrease in item validity from one class to the other and change in mean item response. Here, too, examination of the scatterplots revealed only the common ellipse; a nonlinear relationship was not uncovered.

An interesting supplementary point arose out of the data involved in the two correlation coefficients just mentioned. The mean of the changes in item validity was −.03 (the standard deviation of the changes being .06). The negative sign attached to the mean reflects a general, slight drop or negative change (i.e., toward or beyond zero) in the item validities from the postadmission class to the preadmission class. How much of this drop was due to the unreliability of the item validity statistic and how much to the difference in situations is not known. The over-all distribution of item validities remained very much the same in the two groups, as one can see from the summary statistics in footnote 3 above.

CONCLUSIONS AND DISCUSSION

With regard to the two problems described earlier, the conclusions of the study are:

1) There was only a very slight tendency for the largest item response changes in the direction indicated as desirable by the sign of the validity coefficient to be associated with items of the highest validity.

2) No association was found between drop in item validity from one class to the other and change in mean item response.

While the results were derived from and apply directly only to the particular instrument involved, it seems reasonable that they might be found characteristic of other similar inventories, self-descrip-

tion forms, and questionnaires in which the items are straightforward and not particularly subtle. Although students applying for college are apparently under some pressure intentionally or unintentionally to give what they consider to be desirable answers to these tests, such changes as occur are related only slightly, if at all, to the item validities. This suggests that the preponderance of student opinion as to which is the favorable end of the scale is correct to some extent but that there is great diversity of opinion in this respect about all the items. The suggestion from the second conclusion is that those items which are changed most from the postadmission to the preadmission situation are not any more adversely affected than are those which are changed least.

There is some comfort in these conclusions for questionnaire and inventory constructors. Even though the intent of this kind of item usually appears very obvious to the item writer, it does not follow that the student applying for college will necessarily be able to see through it. In his general effort to paint an attractive picture of himself the student seems rather unable, at least in the situation discussed here, to detect consistently the more valid items so that he may 'put his best foot forward' on these more important items especially. The emphases he injects into his questionnaire replies have little relationship to the validities of the items.

Another implication should be mentioned. The findings of the study raise the possibility that we need not necessarily discard a questionnaire or inventory for use in selection just because (1) the items in the instrument are not subtle and because (2) the normal item response for an applicant group is not the same as the normal response for an admitted group. Responses to the identical items may be quite different for the two groups, but many of the same items may be operative in differentiating among the members within each group. Where it is feasible, any instrument that appears to measure an important kind of behavior might well be worth trying even though the respondents are seeking a goal and might alter their responses accordingly. The empirical evidence will provide the evaluation required and proper use of the scores may still produce valuable information for selection purposes.

SUMMARY

An attitude-interest inventory was administered to two successive freshman classes at a women's college, in one case shortly after

the students had been admitted to college and in the other case as a part of the application procedure for admission. The average response change from the postadmission class to the preadmission class was computed for each item and plotted against the item validity for predicting freshman-grade performance relative to scholastic aptitude. The average item response change was also plotted against the change in item validity from one class to the other. There was only a very slight tendency for the largest item response changes in the direction indicated as desirable by the sign of the item validity coefficient to be associated with items of highest validity. No association was found between drop in item validity from one class to the other and change in mean item response.

REFERENCES

1) G. K. Bennett and H. P. Gordon. "Personality test scores and success in the field of nursing." *J. appl. Psychol.*, 1944, 28, 267–278.

2) H. L. Hollingworth. *The Psychology of Functional Neuroses.* New York: Appleton, 1920, Ch. VIII.

3) R. C. Myers. "A study of rationalization." *J. educ. Psychol.*, 1950, 41, 149–160.

4) R. C. Myers and D. G. Schultz. "Predicting academic achievement with a new attitude-interest questionnaire-I." *Educ. psychol. Measmt.*, 1950, 10, 654–663.

5) F. L. Ruch. "Ability of adults to fake desirable responses on two personality self-inventories and an attempt to develop a 'lie detector' key." *Psychol. Bull.*, 1945, 42, 539–540.

6) E. L. Schott. "Personality tests in clinical practice." *J. abnorm. soc. Psychol.*, 1937, 32, 236–239.

7) D. G. Schultz and B. F. Green, Jr. "Predicting academic achievement with a new attitude-interest questionnaire II." *Educ. psychol. Measmt.*, 1953, 13, 54–64.

8) R. M. Simpson. "Self-rating of prisoners compared with that of college students." *J. soc. Psychol.*, 1933, 4, 464–467.

9) N. B. Van Wagenen. *The Stability of Self-description Tests of Personality Adjustment.* New York: Author, 1935, pp. 46–47.

INFLUENCE OF SOCIAL CLASS UPON PERFORMANCE ON THE DRAW-A-MAN TEST

JOSEPH H. BRITTON

The Pennsylvania State College

Sociologists and cultural anthropologists are stressing the importance of the location of an individual in the social-class structure of his community as a basic determiner of a large proportion of the cultural and environmental experiences he may have (*14*, *15*). By limiting the environmental experiences in socialization processes, a social-class group tends to restrict the individual's learning to a particular social pattern of behaving and believing, a pattern in many ways distinct from that of persons belonging to (or identifying with) other social classes (*4*, *12*). The differences in cultural experience, together with an individual's cultural motivation, may be expected to influence the content of what a child will learn in school, how he will feel about his teacher, his peers, his school experience in general (*3*, *13*, *17*). These cultural differences may be expected to affect also the responses of children on intelligence tests (*2*). The problem of this investigation is to discover the effect of social class membership upon performance on certain standardized group tests of intelligence and upon performance on the Draw-a-Man Test (*1*).

Research on the relation of environmental-cultural factors and intelligence has been carried on by many investigators. Their work is reviewed by Kenneth Eells (*5*), as well as by Virgil Herrick (*11*). "The research is close to unanimous," writes Herrick, "in showing that there are significant differences in intelligence-test performance of children and youth from different socio-economic backgrounds, with children from the higher levels always securing the higher intelligence test scores." (p. 12).

A number of studies have shown the effect of environment upon the responses of children on the Draw-a-Man Test (*8*); among these is the research of Havighurst and his associates (*9*) concerning the performance of three hundred and twenty-five Indian children of the Sioux, Hopi, Zuni, Zia, Navaho, and Papago tribes in nine communities. They found that Indian children from these tribes

did better than white children on the Draw-a-Man Test, that Indian boys from the Pueblo groups did better than girls on the test, the evidence pointing strongly to the conclusion that environment affects the performance of children on the Draw-a-Man Test. To study the effect of social class environment on performance of white children on the Draw-a-Man Test is the problem of this investigation.

PROCEDURE

Subjects for the study include one hundred and two boys and one hundred and thirty girls whose social status in their midwestern community had been determined by using Warner's Index of Status Characteristics (*16*). Warner's index is based upon ratings, on seven-point scales, of four different status characteristics—occupation, source of income, type of house lived in, and dwelling area in the community. These four ratings are combined into a single index (ISC) by means of a weighting scheme; in this particular investigation, a part of a study of the University of Chicago (*6*), education of the parents was substituted for the source of income. The ISC has been shown to be closely related to social class placement as determined by actual social participation and social reputation in the community.

Standardized verbal and nonverbal group tests of intelligence (Henmon-Nelson, Otis Alpha Verbal, Otis Alpha Non-Verbal, Kuhlmann-Anderson) were administered to these nine-, ten-, and eleven-year-olds in their public and parochial school classrooms. The following year the children were asked to "draw the best man you can," and each of these drawings was scored twice on the Goodenough scale (*7*) by two assistants.[1]

The comparisons between high and low socio-economic groups of boys and girls on the mean IQ's and their standard deviations were made for the five tests mentioned above. Tests of significance were made between status-sex groups on the Draw-a-Man Test. Then product-moment correlations were calculated between IQ's on the various tests and the ISC's.

---

[1] The aid of Miss Helen Coburn and Mr. David Twain, graduate assistants, Department of Psychology, The Pennsylvania State College, is gratefully acknowledged.

TABLE 1.—MEAN INTELLIGENCE QUOTIENTS AND STANDARD DEVIATIONS ON INTELLIGENCE TESTS FOR HIGH AND LOW SOCIO-ECONOMIC GROUPS OF 9-, 10-, 11-YEAR OLD BOYS AND GIRLS

Test	High SE Group*						Low SE Group†					
	Boy			Girls			Boys			Girls		
	N	Mean IQ	SD	N	Mean IQ	SD	N	Mean IQ	SD	N	Mean IQ	SD
Henmon-Nelson	52	115.0	14.4	67	114.1	16.3	41	94.2	13.4	55	98.2	15.2
Otis Alpha Verbal	54	106.9	9.4	66	102.8	10.4	43	96.3	9.9	56	93.8	10.0
Otis Alpha Non-Verbal	54	104.3	10.2	66	105.3	9.9	42	93.2	8.9	56	94.5	10.3
Kuhlmann-Anderson	53	107.5	8.7	69	108.6	10.1	41	93.0	10.5	58	96.8	11.3
Draw-a-Man‡	55	81.0	17.6	69	89.4	15.2	43	82.1	13.6	58	83.2	12.6

* ISC's 5–15 inclusive.

† ISC's 21–27 inclusive.

‡ Calculations were made to determine the statistical significance of differences between status-sex groups of mean IQ's on the Draw-a-Man test; all three differences involving the high status girls were significant.

RESULTS AND DISCUSSION

The results of the comparisons between high and low socio-economic groups are shown in Table 1. The high status group consists primarily of upper-middle and upper class boys and girls, and the low status group contains lower class children. It will be observed that both the high socio-economic group and the low socio-economic group were further divided into boys and girls. For both boys and girls of the high status group mean IQ's on the Henmon-Nelson, Otis Verbal, Otis Non-Verbal, and the Kuhlmann-Anderson are all above 100, the highest being the average IQ for the boys, 115.0. In contrast, the average IQ's for boys and girls of the low-socio-economic group all remain under 100, the highest in this case being the girls' mean Henmon-Nelson IQ 98.2. These results are very similar to those reported by Eells, of course, since our study population was a part of his. He found that the differences between his status groups, ranging from eight IQ points to twenty-three IQ points, were statistically significant. Our results are interpreted here to mean that clearly there are differences between the low and high socio-economic groups as to their performance on these widely-used intelligence tests. Moreover, as commonly observed, the differences favor the higher status group.

With regard to the mean Draw-a-Man IQ's for all four groups compared, they were definitely lower than the others; they ranged from 81.0 for the high status boys to 89.4 for the high status girls. Just why these average IQ's should be so low as compared to those of the other standardized tests is uncertain. There is, of course, the possibility of the pupils' not being motivated adequately enough to score well on the Goodenough; in gaining rapport in the classrooms the testers who aided us by administering the tests in the larger research project (*6*) endeavored to put the pupils at ease. The result may have been that they relaxed their efforts and felt that this was not a test but was designed simply to pass the time. Consequently, the results may be spuriously low. It should be pointed out, however, that in some instances the variability of the groups on the Draw-a-Man test, as shown by the standard deviations, is larger than on other tests.

Tests of the significance of differences between mean Draw-a-Man IQ's revealed that all those involving the high status girls were beyond chance expectations, those girls' average IQ being significantly higher than others. In Havighurst's investigation (*9*) involving the Indian children, a contrasting sample of Midwest white children was obtained; within that group there was a slight superiority of girls over boys on the Draw-a-Man Test, though the difference, they say, may have been due to chance.

Further average IQ's on the Draw-a-Man were calculated and are shown in Table 2. The difference between the mean IQ of the total high status group and that of the total low status group was not statistically significant; the difference between the mean IQ's of all boys and all girls was significant at the one per cent level.

Table 3 shows the results of correlation analysis of the IQ scores on the various tests with the Index of Status Characteristics. The

TABLE 2.—MEAN DRAW-A-MAN IQ'S AND SD'S SELECTED GROUPS

Group	N	Mean IQ	SD
High S-E group	124	85.6	16.5
Low S-E group	101	82.7	13.1
All boys*	98	81.5	15.5
All girls*	127	86.5	14.5

* The difference between mean IQ's of all boys and all girls was statistically significant.

TABLE 3.—PRODUCT-MOMENT CORRELATION COEFFICIENTS OF ISC'S WITH IQ'S ON VARIOUS TESTS

Test	N	r and PE_r
Henmon-Nelson	222	.51 ± .04
Otis Alpha Verbal	226	.42 ± .04
Otis Alpha Non-Verbal	223	.50 ± .04
Kuhlmann-Anderson	228	.53 ± .03
Goodenough Draw-a-Man	232	.11 ± .04

four coefficients involving the Henmon-Nelson, Otis Verbal, Otis Non-Verbal, and the Kuhlmann-Anderson are all moderate, *r*'s ranging from .42 to .53. Each *r* is statistically significant, being larger than four times its probable error. The higher the social status, as measured by the ISC, the higher is the intelligence, as measured by these IQ tests. This tendency is in conformity with other investigations concerning the relationship between intelligence and socio-economic background.

The coefficient showing the correlation between the response on the Goodenough Draw-a-Man Test and the ISC is .11 with its probable error indicating that this small degree of relationship is unreliable. In contrast, then, with the previous tests, this shows little or no relationship between the Draw-a-Man IQ and socio-economic status. Performance on the Goodenough scale does not appear to be influenced by the children's membership in a particular social class group. This finding may be substantiated by data in a study of ten-year-old white children by Havighurst and Janke (*10*), whose data do not lend themselves to rigid comparison with ours. In the midwestern community they studied, some differences were revealed between the status groups on the Goodenough scale. In the small samples they compared on this particular test, the mean IQ for lower-lower class children was 91; for lower-middle class children the mean IQ was 107, the difference between these socio-economic groups being 16.

One of the assumptions which test constructors utilize is that all subjects taking the test must be equally familiar or equally unfamiliar with the test materials. In fact, Goodenough's selection of a man as the object of this test was on the basis that all children (presumably all children in American cultures) have had approximately equal opportunities to observe the details of the human

figure on which scoring was standardized, so that resulting variations could be said to be due to discrepancies among children in their capacity to make use of such opportunities, rather than to their not having had such opportunities.

A legitimate hypothesis might be to the effect, however, that on the Draw-a-Man Test children of higher socio-economic status would perform somewhat better than lower class children. This we felt might be true because the higher status group would probably have had more experiences with drawing materials and with pictures of people in books and magazines; their relationship with school teachers would probably have been more positive, and the degree of their motivation to do well on the test higher than that of the lower class children. Moreover, the attention most middle and upper class parents give to details of clothing, personal neatness, etc., might be expected to influence their children's awareness of specific details which in turn might increase appreciably their scores on the Goodenough.

This difference in performance by social class groups, we might further hypothesize, would be less than with the other tests analyzed because of the almost complete non-verbal nature of the Goodenough and its administration, as compared with tests using verbal and picture-type items, which have been shown to favor often-times the upper status groups.

It would appear that on the basis of our data, social class environment has not affected to any significant extent the responses on the Draw-a-Man Test; but the differences which are apparent between the correlation coefficients of the Henmon-Nelson, Otis Verbal, Otis Non-Verbal, the Kuhlmann-Anderson, and the Draw-a-Man do appear to substantiate the latter hypothesis on the relative influence of social class. It is believed that the nature of the test materials themselves is the largest contributing factor. A significant difference between performance of low status groups and higher status groups on the Draw-a-Man was not revealed.

SUMMARY AND CONCLUSIONS

In this paper we have summarized findings of a study of the performance of some elementary school children on certain widely used intelligence tests. Having obtained their families' socio-economic status ratings in their midwestern community, by means of Warner's Index of Status Characteristics, we were able to compare

their IQ scores by social class. Further, having administered and scored Goodenough Draw-a-Man Tests on these children, we were able to show the relationship which existed between performance on the Draw-a-Man Test and social status. This association proved to be negligible; the relationships between social status and IQ's on the Henmon-Nelson, Otis Alpha Verbal, Otis Alpha Non-Verbal, and the Kuhlmann-Anderson were moderate but statistically significant.

The widespread use of intelligence tests as selectors of human abilities—as instruments to help educational counselors in guiding students and to aid industrial and military leaders in selecting personnel—is well-known. We submit that the optimum utilization of human resources, actual and potential, in all socio-economic groups is an obligation. The search for measures of intelligence which resolve social-class differentials must, therefore, be continued; questions of motivational factors which influence the performance of children on tests should receive concerted attention along with research on the content of tests themselves.

REFERENCES

1) J. H. Britton. "Influence of social class upon performance on the Draw-a-Man Test." *Amer. Psychologist*, 1952, 7, 304. (Abstract)

2) A. Davis. *Social-class Influences upon Learning*. Cambridge: Harvard University Press, 1949.

3) ———— "Socialization and adolescent personality." In N. B. Henry (Ed.). *Forty-Third Yearbook of the National Society for the Study of Education*. Part 1, *Adolescence*. Chicago: Department of Education, University of Chicago, 1944.

4) A. Davis and R. J. Havighurst. *Father of the Man: How Your Child Gets His Personality*. Boston: Houghton Mifflin, 1947.

5) K. W. Eells. *Social-status Factors in Intelligence-test Items*. Unpublished doctor's dissertation, University of Chicago, 1948.

6) K. Eells, A. Davis, R. J. Havighurst, V. E. Herrick, and R. Tyler. *Intelligence and Cultural Differences: A Study of Cultural Learning and Problem-solving*. Chicago: University of Chicago Press, 1951.

7) Florence L. Goodenough. *Measurement of Intelligence by Drawings*. Yonkers, N. Y.: World Book Company, 1926.

8) Florence L. Goodenough and D. B. Harris. "Studies in the psychology of children's drawings: II 1928–1949." *Psychol. Bull.*, 1950, 47, 369–433.

9) R. J. Havighurst, Minna Gunther, and Inez Pratt. "Environment and the Draw-a-Man Test: the performance of Indian children." *J. abnorm. soc. Psychol.*, 1946, 41, 50–63.

10) R. J. Havighurst and Leota L. Janke. "Relations between ability

and social status in a midwestern community. I. Ten-year-old children." *J. educ. Psychol.*, 1944, 35, 357–368.

11) V. E. Herrick. "What is already known about the relation of the IQ to cultural background?" In K. Eells, *et al. Intelligence and Cultural Differences: A Study of Cultural Learning and Problem-solving*. Chicago: University of Chicago Press, 1951.

12) W. L. Warner and Associates. *Democracy in Jonesville: A Study of Quality and Inequality*. New York: Harper & Brothers, 1949.

13) W. L. Warner, R. J. Havighurst, and M. L. Loeb. *Who Shall Be Educated?—the Challenge of Unequal Opportunities*. New York: Harper & Brothers, 1944.

14) W. L. Warner and P. S. Lunt. *The Social Life of a Modern Community*. Vol. 1. *Yankee City Series*. New Haven: Yale University Press, 1941.

15) ———— *The Status System of a Modern Community*. Vol. 2. *Yankee City Series*. New Haven: Yale University Press, 1942.

16) W. L. Warner, Marchia Meeker and K. Eells. *Social Class in America: A Manual of Procedure for Measurement of Social Status*. Chicago: Science Research Associates, 1949.

17) *Progressive Educ.* 1950, 63, No. 4 (whole issue).

CLASSROOM ENVIRONMENT AND PUPIL WELFARE

NEWELL C. KEPHART

Purdue University

and

WILLIAM FLOYD

West Lafayette (Indiana) Public Schools

Recently public school systems have given considerable attention to the physical environment of the school classroom. Traditional color schemes with their preponderance of dull shades of gray and buff have given way to brighter and more cheerful decorative motifs. Woodwork, floors and furniture have changed from traditional dark oak to lighter shades of sand or bleached finishes. Rigid desks in straight rows have been replaced by movable desk and chair combinations which can be adjusted to the physical requirements of the pupil. Control of both natural and artificial light is distributing the quantity of available illumination more evenly and increasing its usefulness for each pupil in the classroom.

In the Morton Elementary School, West Lafayette, Indiana, a series of research studies was undertaken to determine the effects of these new classroom designs upon the welfare of pupils who spent a school year in this new environment. The room design used followed that of the 'coördinated classroom' recommended by Harmon.

Two classrooms were refinished in accordance with these more modern developments. Three major changes were undertaken. First, the distribution of light was rearranged so that the quantity and quality of light was distributed equally, as nearly as possible, over all portions of the room. Care was taken to see that at any seat in the room adequate light was available for the critical recognition of visual detail and the full and free performance of visual tasks, bearing in mind that such visual tasks are not confined to a single plane but are three dimensional in character. An optimum distribution of light was obtained by installing semi-trans-

parent baffles at an angle to the upper half of the windows in such a manner that two thirds of the light entering through the window was reflected onto the ceiling and one third was transmitted in the normal manner. The baffles were set at such an angle that the reflected light from the ceiling was directed to the normally dark sides of the room furthest from the windows. The ceiling and an eighteen-inch strip at the top of the walls were painted white to aid in this distribution of light. Natural light was supplemented by artificial light also designed to be distributed equally over all parts of the room. For this purpose florescent fixtures were installed in reflectors so designed that the light was evenly distributed.

The second consideration in the refinishing was the amount of contrast between the visual task presented to the child and its surrounding. It is desirable that this contrast be reduced markedly from that which occurs in the normal classroom and preferably be not greater than three to one. Since most schoolroom taks involve working with white paper, the desks in the experimental room were finished in a light wood finish which reduced markedly the contrast between the paper on which the child was working and the surrounding desk tops. For the same reason the finish on the floor was light. Dark woodwork was replaced by light finish and pastel shades were used in the painting of the walls. Black chalkboards were replaced by yellow green chalkboards to further reduce undesirable contrast.

The third change involved the furniture in the room. It is desirable that such furniture permit freedom of movement to the child and be adjustable both to the physical size of the child and to the task at hand. The American Seating Company ten-twenty desk was chosen and was used to replace the previous furniture. These desks were not placed in the customary straight rows but were arranged in slightly curved rows so that as nearly as possible a fifty degree angle between the plane of the windows and the child's work was maintained. Such an arrangement minimizes the undesirable shadows which fall on a pupil's desk.

Two additional rooms identical to the experimental rooms, as to location on the side of the building, number of windows, etc., were chosen as control rooms. For each maintenance operation which was performed in the experimental classroom, a similar operation was performed in the control classroom. When the experimental classrooms were painted, the control classrooms were

also painted but the traditional color scheme was followed. When new furniture was provided in the experimental classrooms; new furniture, but of a more traditional design, was provided in the control classrooms. Similarly, with all other necessary alterations in the experimental classrooms, a similar amount of work, material, and money was put into the control classrooms. As a result the control rooms were just as new, just as clean, and just as completely refinished as were the experimental rooms. The only difference was that the scheme followed in refinishing the control rooms was traditional whereas the scheme followed in refinishing the experimental rooms was that of the coördinated classroom.

One experimental room and one control room were designed to be occupied by fourth-grade classes. The second experimental room and the second control room were designed to be occupied by fifth-grade classes. At the beginning of the school year, all students enrolling in the fourth grade were assigned at random to either an experimental or a control room. Similarly all students enrolling in the fifth grade were assigned at random to either an experimental or a control room. By this procedure, two randomly selected groups were obtained at each grade level, one assigned to the experimental room and one assigned to a control room. This random selection resulted in two groups in each grade which did not differ significantly in educational achievement level or IQ.

No attempt was made in this series of experiments to evaluate all of the possible effects of the experimental room environment upon school children. Similarly no attempt was made to investigate all of the hypotheses which have been advanced in support of the coördinated classroom. Rather, three general areas of investigation were chosen which, it was thought, were important to the welfare of school children. The three areas selected were school achievement, posture, and frequency of nervous habits. An experimental design was developed in each of these three areas which it was felt would throw light upon the question of differences between the experimental and control groups. The results of these three experiments are reported here.

EDUCATIONAL ACHIEVEMENT

In investigating the variable of school achievement, scores on the Stanford Achievement tests, intermediate examination and the Henmon-Nelson test of mental ability were used. Only the two

fifth-grade classes could be used for this experiment since the tests administered at the fourth-grade level were not comparable. These results are therefore based on twenty-three children in the fifth-grade experimental class and twenty-four children in the fifth-grade control classroom.

For each child the accomplishment quotient was computed. This quotient evaluates educational achievement in terms of both chronological age and mental ability and, therefore, represents achievement corrected for the mental ability level of the pupil. These quotients were then averaged for each of the two classes separately. For the experimental group, the mean accomplishment quotient was 92.1. For the control group, the mean accomplishment quotient was 85.4. The difference between the two groups was 6.7 in favor of the experimental classroom. This difference was found to be statistically significant with a *t*-value of 3.48.

From these results it is seen that the educational achievement as measured by the accomplishment quotient was greater in the class which spent the school year in the experimental environment than in the case of the class which spent the school year in the control environment. The difference between these two classes was significant at less than the one per cent level of confidence.

POSTURE

In investigating postural differences between the two groups of children, only one aspect of total posture, head-tilt, was considered. A head-tilt is defined as a tendency on the part of the child to hold his head tilted noticeably to either one side or the other when he is standing erect and performing a visual task which is located at eye level and twenty feet in front of him.

To measure this variable, a photographic technique was used. A 35 mm. camera was mounted on a tripod and placed directly in front of the child but below his eye level. Across the room at a distance of twenty feet a Snellen visual acuity test chart was mounted on the wall at the child's eye level. The child was brought into the photographic room and was shown a circle eighteen inches in diameter drawn upon the floor with chalk. The Snellen chart across the room was pointed out to him, and he was instructed to stand in the circle and read the letters on the chart. No further instruction was given as it was desired not to influence the in-

dividual's posture by the instructions. When the child had read down the Snellen chart to the point where the letters became difficult for him to recognize, a full length front view photograph of him was taken.

When these photographs had been developed and printed, judges were requested to determine whether or not a head-tilt was present for each child. The instructions to the judges were to identify the midline of the body by observing the sternal notch and a point midway between the knees. They were to observe the midline of the face by determining a point midway between the eyes and the point of the chin. If the midline of the body and the midline of the face coincided, the judge was to indicate that a head-tilt was not present. If the two lines formed an angle, he was to indicate that a head-tilt was present.

From an observation of the photographs, it was apparent that the measurements represented a bimodal curve. If a head-tilt was present, it was present to a marked degree and it was clearly recognizable. There were extremely few cases in which the head tilted so slightly as to leave doubt in the minds of the judges. There appeared to be very few cases of slight or doubtful head-tilt. For this reason the measurement of these photographs by the method of judgment or rating and the establishment of a bi-category variable appeared to be justified and was followed in computing further results.

The reliability of the rating method for determining posture described above was investigated statistically. In this investigation, two problems arose. The first problem was the reliability with which different raters would evaluate the same set of photographs. To test this aspect of reliability, two raters were asked to judge the same set of photographs. Their results were then compared using the tetrachoric r as a measure of agreement. This procedure resulted in an r of .97.

The second reliability problem involves the degree to which the postural characteristics revealed by the photographs would remain constant from day to day. To test this problem, a group of children were photographed on one day using the procedure described above. On the following day the same children were photographed again using the same identical procedure. A rater was asked to judge both sets of photographs according to the procedure outlined above. The result of his judgment on these two sets of

data were then compared, using the tetrachoric *r* as the measure of agreement. This procedure resulted in a coefficient of .81, which indicates that the postural characteristics revealed by this photographing technique are relatively constant from day to day and consequently indicate an habitual posture which persists and tends to recur whenever the child is performing a task similar to that represented by the present experimental situation.

When the reliability of the photographic technique and the scoring method had been established, the number of children with head-tilt was determined for all groups in both grades and in both environments. There were thus a total of forty-nine children who had spent one full year in the control environment. Of these control children, forty-nine per cent displayed a head-tilt when measurements were taken in May. There were also forty-nine children who had spent one school year in the experimental environment. Of these experimental children, only twenty-seven per cent showed a head-tilt when measurements were taken in May. There was thus a difference of twenty-two per cent between the groups. This difference was statistically significant, the *t* value being 2.18. It would appear that the experimental environment has a less detrimental effect upon the habitual posture of the child as indicated by the presence of head tilts.

FREQUENCY OF EXTRANEOUS MOVEMENTS

Subjective observation of the two types of classroom environment involved in these studies had suggested that children in the control classrooms might show a higher frequency of extraneous movements or nervous habits. In order to investigate this possibility, the technique developed by Olsen was used. This technique consisted of an observation of a timed sample of behavior. Olsen had constructed a list of the most common nervous habits displayed by young children. The technique consisted of observing a child for a predetermined period of time and noting for each habit in the list, the number of times which it occurred during the observation period.

Using Olsen's list of nervous habits, similar observations were made in the four classrooms involved in the present study. In September, and again in May, an observer went into the classroom. With Olsen's list before him, he observed the behavior of the first child in the room for a two minute period. Whenever one of

the habits on Olsen's list was observed, a tally mark was made on the scoring sheet. At the end of the two minutes, the observer transferred his attention to the next child whom he observed for a similar two-minute period. These observations were continued until all of the children in the room had been observed. The total number of tally marks on each score sheet was determined and this figure was considered the raw score on the nervous habits inventory for that child. The score thus represented the number of times that activities listed by Olsen occurred during a two-minute observational period. Each of the four classrooms was treated in the same manner.

In order to determine the reliability of this technique, two observers were sent into each of two rooms and made observations according to the procedure outlined above simultaneously on each child in the room. The results of correlating the ratings by each of the two raters was a reliability coefficient of .93.

In order to determine any difference between the two types of classrooms, the results were treated separately for the control rooms and the experimental rooms. In the control rooms, the average frequency of nervous habits in the September observations were computed. A similar average was computed for the May observations. The difference between these two averages was noted. This difference was found to be $-.63$, indicating that the average frequency of nervous habits was slightly higher in May than in September for this group. This difference however, was not statistically significant as the t value was only .95.

Similar averages were obtained for the experimental groups and the difference computed. The difference in these experimental groups was found to be 1.52, indicating that the frequency of nervous habits was less in May than in October. These experimental group differences were found to be statistically significant, having a t value of 2.71.

It is thus seen that the children who spent a school year in the control type classroom showed no change in the number of nervous habits which they displayed in a two-minute time period. On the other hand, the children who spent this school year in the experimental classroom environment showed a decrease in the number of nervous habits which they displayed in a similar two-minute time period.

It appeared that the experimental environment was having some

effect upon the number of extraneous movements or nervous habits which the children displayed. If it can be argued that these nervous habits are an indication of the level of tension under which the individual is working, then it can be similarly argued that this tension is less in the experimental type classroom than in the control type classroom.

SUMMARY AND CONCLUSION

This study attempted to evaluate the effect upon fourth- and fifth-grade school children of a school year spent in an experimental classroom. The experimental classroom used was that recommended by Harmon and known as the coördinated classroom. A group of children who spent a school year in this type of classroom was compared with a similar group who spent the same school year in a traditional type classroom environment. Three areas of development were studied: school achievement, posture, and frequency of extraneous movements or nervous habits. It was found that the group who spent the school year in the experimental environment was superior to the group who spent the school year in the control environment in school achievement as measured by the accomplishment quotient, in the posture of the head in relation to the body (a smaller percentage of head-tilts), and in extraneous behavior patterns (a lesser frequency of extraneous movements or nervous habits). In the areas investigated in this study, it is concluded that the newer type classroom environment represented in this case by the coördinated classroom can have a beneficial effect upon school children who spend a considerable portion of their time in this environment during the school year.

BOOK REVIEWS

ROGER G. BARKER, BEATRICE A. WRIGHT, LEE MEYERSON, AND MOLLIE R. GONICK. *Adjustment to Physical Handicap and Illness: A Survey of the Social Psychology of Physique and Disability*. New York: Social Science Research Council, (Bulletin 55, Revised), 1953, pp. 440.

The social psychology of physique and disability is a problem that has obtained in ever increasing extent the interest of both research people and workers in the rehabilitation field. Bulletin 55, the original volume of the Social Science Research Council which surveyed this topic has influenced the direction as well as the amount of interest and work on the subject. This appeared in 1946. Enough changes have taken place since, in the opinion of the investigators, to warrant a revised edition. The pattern of organization, selection and sequence of topic considerations is the same as the first. The findings are reported in six chapters with the same captions, and a bibliography follows. There are changes, and the changes are largely in terms of what the investigators consider to be changing trends calling for a reappraisal to serve as a guide for still further research and still more realistic and needed applications.

The outstanding change in this revised edition is in the shift away from emphasis on statistical facts and correlations to a larger emphasis on the use of human documents. In the revised edition there is a definite improvement in literary style. The revision is more than merely a repair job. It is that, but it is more. The changes made are sometimes just in rephrasing of a word, reconstruction of a phrase, and at times a paragraph. The most marked changes are in the first two chapters: The introductory chapter on methodological problems of somatopsychological research is described in some detail with a much more substantial perspective for comprehensive orientation for understanding all the problems than is done in the original volume. Also Chapter 2 concerned with normal variation in physique is presented in a fuller fashion. To a small extent, such differences as appear between the original edition and the present revised edition are similar in nature to those found in the two volumes of Middletown by the Lynds: *Middletown* full of statistics

and with the omission of human factors and *Middletown in Transition* beginning the book with the Ball family and describing the people behind the scenes. The changes here are not that drastic and marked, but they are in the same direction of increased significance of research and applications rather than a mere listing and tabulation. The chief changes are in a broader comprehensive orientation in defining the field both in increased clarity of concepts used and an increased searching for theories of soma-psyche relations, so there's a more frequent searching for and explanation of likely explanatory hypotheses. Also revealed in the more recent studies is improved technical adequacy of some of the research. These changes for the better are stressed and influence the revised edition. But they also do one other thing—they bring to the focus of attention some of the urgent needs that still remain in the field. Everything is considered related to the problem they are concerned with that, in their way of defining it, is "Concerned with those variations in physique that affect the psychological situation of a person by influencing the effectiveness of his body as a tool for actions or by serving as a stimulus to himself or others." The list of physical variations which can have significance as listed by them are eighteen in number and include size, race, age, sex, cosmetic defect, muscular strength, motor ability, speech defects, auditory impairments, visual impairments, tuberculosis, heart disease, diabetes mellitus, rheumatism, leprosy, cancer, orthopedic disabilities, acute illness. Also included in their consideration are some psychoanalytic contributions such as Schilders in consideration of the body image in relationship to the self-concept. And the investigators call attention to one practical phase of the problem; namely, the urgent need for investigating the effectiveness of forms of coping with the problems of disabled workers.

All in all, the investigators made a competent survey of the literature in their first effort, and the people who are in the field, either in the field of rehabilitating and educating the disabled or social scientists interested in the followup of studies of social psychology of physique and disability, will find this volume an even more satisfactory frame of reference for orientation in the field than was the first volume.

Psychological Service Center H. MELTZER

MARJORIE B. SMILEY. *Intergroup Education and the American College*. New York: Bureau of Publications, Teachers College, Columbia University, 1952, pp. 212, $4.25.

The foundation for this latest volume of the Teachers College Studies in Education is provided by a detailed and intelligent analysis of the intergroup education programs reported by fifty-six colleges and universities in response to questionnaires submitted by the author. Each chapter presents the philosophy of a certain aspect of intergroup education, a summary of policies and practices of the fifty-six institutions in regard to this aspect, and recommendations for planning a curriculum to deal with this aspect. The organization should be convenient for individual college instructors or committees charged with the task of working out a curriculum concerned with desirable attitudes, values and skills in intergroup relations. To such committees and teachers the book can be recommended. The emphasis on curriculum development rather than methodology makes it of limited value to most workers in the social sciences.

Neither the author's thoroughness nor her command of the subject and related areas will be challenged by many readers of this volume. Profuse quotations and citations of authority plus a completely-thought-through concept of intergroup education leave no doubt of the writer's familiarity with the subject. However, this style of writing is probably also the major weakness of the volume. Perhaps the author felt the necessity to reinforce her arguments with the statements of authorities, because she is aware that some college teachers might not see the need for intergroup education. Whatever the reason, the style makes for low readability from the standpoints of both interest and difficulty, and may discourage some readers.

University of Oklahoma WILLIAM ELLER

BERNARD H. HALL, MARY GANGEMI, V. L. NORRIS, VIVIENNE HUTCHENS VAIL AND GORDON SAWATSKY. *Psychiatric Aide Education*. New York: Grune and Stratton, 1952, pp. 168.

Psychiatric Aide Education is a report of a seven-year experiment of a training program carried out by the Menninger Foundation in coöperation with the Topeka State Hospital. A grant from the

Rockefeller Foundation supplied the necessary funds. The school opened officially on October 3, 1949, and closed on June 30, 1952. A total of eighty-five students were matriculated; sixty of them graduated. The story of this project includes a report concerning the field of psychiatric aide—its growth and its needs; recruitment and selection of students; an evaluation of the curriculum; a story concerning its graduates; and an evaluative chapter on recent developments in psychiatric aide education. Also included is a fair sized appendix, which contains a bibliography and forms used or relevant in considering the problem of psychiatric aide education. The forms include application for admission, course outlines, and course evaluations.

All in all, it's an interesting story written by the staff members. Dr. Karl Menninger in his preface describes the significance of this project as well as his reaction to the obstacles experienced because of prejudice and blind spots in unexpected places. This story should be of interest to hospital administrators and those concerned with the problem of psychiatric aide education.

Psychological Service Center H. MELTZER

EDWIN FORBES TAIT. *Textbook of Refraction.* Philadelphia: W. B. Saunders Company, 1951, pp. 418.

This book on refraction by a professor of ophthalmology at Temple University is primarily intended for a student group preparing for the practice in the field of eye work. Accordingly, the author has sensibily enough selected to write the book in straightforward, at times seemingly dogmatic form, using references that are pertinent and readily available. Also included by him are books and papers suggested as further reading on the more original material that can be used by more ambitious students or the more curious and inquiring readers.

The text as written presupposes a background of fair knowledge of the anatomy and physiology of the eyes and the nervous system as well as an acquaintance with the principal methods of geometric and physiological optics. The early chapters contain discussions of the basic material necessary for the study of refractive error and ocular motor anomalies. They also contain a short, elementary presentation of the more important concepts which are discussed in greater detail later in the volume. There is a chapter on vision

that continues with a consideration of preliminary concepts for understanding problems in the field and then chapters are outlined that are of anatomical structure or of functions related to the field. Illustrative of topics included are: ocular reflexes; anomalies and variations of ocular accommodation; myopia; astigmia; presbyopia and aphakia; binocular fusion, and principles and methods in heterophoria.

In all, this is a carefully prepared volume that should serve well to do what it proposes to; namely, serve students in the field as well as neurologists and psychologists who are interested in having an authoritative volume on ocular refraction.

Psychological Service Center H. MELTZER